Investigating the genetic predisposition for familial haematological malignancies

By

Dr Elizabeth Mary Tegg MBBS, FRCPA

Abbreviations

ABL1 v-abl Abelson murine leukaemia viral oncogene homolog 1

ALL Acute lymphoblastic leukaemia

AML Acute myeloid leukaemia

AML M1 Acute myeloid leukaemia with minimal differentiation

APML Acute promyelocytic leukaemia

ATM Ataxia telangiectasia mutated

BCR Breakpoint cluster region

BMI-1 Polycomb complex protein BMI-1

BRIP1 BRCA1 interacting protein C-terminal helicase 1

CCND1 Cyclin D1

CDH24 cadherin 24, type 2

cDNA coding DNA

CHEK2 CHK2 checkpoint homolog

CHL Classical Hodgkin lymphoma

CLL Chronic lymphocytic leukaemia

CML Chronic myeloid leukaemia

CMML Chronic myelomonocytic leukaemia.

CNV Copy number variation

DLBCL Diffuse large B cell lymphoma

DNA Deoxyribonucleic acid

EBI2 G protein-coupled receptor 183

ET Essential thrombocythemia

FFPET Formalin fixed paraffin embedded tissue

FGR1 Gardner-Rasheed feline sarcoma viral oncogene homolog

FL Follicular lymphoma

FLT3 fms-related tyrosine kinase 3

GATA-2 GATA binding protein 2

GFI-1 Growth factor independent 1 transcription repressor

GPR123 G protein-coupled receptor 123

GRINL1B glutamate receptor, ionotropic, N-methyl D-aspartate-like 1B

HCL Hairy cell leukaemia

HL Hodgkin lymphoma

HL LD Lymphocyte depleted CHL

HLLR Lymphocyte-rich CHL

HLMC Mixed cellularity CHL

HLNS Nodular sclerosis CHL

HM Haematological malignancy

HSC Haematopoietic stem cell

KNDC1 kinase non-catalytic C-lobe domain containing 1

LMO1 Rhombotin-1

LPD Lymphoproliferative disorder

MALT Mucosal associated lymphoid tissue

MBCN Mature B cell neoplasm

MBL Monoclonal B cell lymphocytosis

MCL Mantle cell lymphoma

MDS Myelodysplastic syndrome

MF Myelofibrosis

MLL mixed lineage leukaemia

MM Multiple myeloma

MPN Myeloproliferative neoplasm

MTCN Mature T cell neoplasm

NHL Non Hodgkins lymphoma

NOS Not otherwise specified

NPM1 nucleophosmin

PALB partner and localizer of BRCA2

PBALL Precursor B lymphoblastic leukaemia/lymphoma

PCLKC cadherin-related family member 2

PDGFRA Platelet-derived growth factor receptor, alpha polypeptide

PDGFRB Platelet-derived growth factor receptor, beta polypeptide

PTALL Precursor T lymphoblastic leukaemia/lymphoma

PTPRT protein tyrosine phosphatase, receptor type, T

PV Polycythemia Vera

RA Refractory anaemia

RAEBI Refractory anaemia with excess blasts type 1

RAEBII Refractory anaemia with excess blasts type II

RARS Refractory anaemia with ring sideroblasts

RCMD Refractory cytopenia with multilineage dysplasia

RHH Royal Hobart Hospital

RUNX1 Runt-related transcript factor 1

RUNX1T1 runt-related transcription factor; translocated to, 1

SCL/TAL1 stem cell leukaemia/T-cell acute lymphoblastic leukaemia 1

SLC15A1 solute carrier family 15, member 1

SLL Small lymphocytic lymphoma

SNP Single nucleotide polymorphism

TEL(ETV6) ets variant 6

UTR untranslated region

WCC White cell count

WGA Whole genome association

WGS Whole genome sequence

WM Waldenstroms macroglobinulaemia

1. Chapter 1: Literature review

1.1. *Introduction*

Haematological malignancies (HMs) include cancer and related disorders of the cells of the blood, bone marrow or lymphatic systems. Currently the most widely used classification system is the World Health Organization classification system of HMs(1). This classification system classifies HMs into 12 broad categories based upon clinical, morphological, immunophenotypic and genetic characteristics of the tumour cells. Approximately 250 new cases of HMs are diagnosed in Tasmania each year(2).

Familial aggregation of HMs has been repeatedly described in the literature(3-24), but the genetic contributors to this disease in most families remains unknown. The only germline mutation found to date is a DAPK1 mutation which was found in a family with chronic lymphocytic leukaemia (CLL)(25). In this family additional down-regulation of *DAPK1* was still needed before CLL developed(25).

Mature haematological cells originate from a common precursor cell (haematopoietic stem cell (HSC)) that has the capacity to differentiate to any of the mature cell types and to self renew(26). It has been established that for a cell to become malignant multiple genetic mutations must occur in the same cell. In familial cancer families it is reported that the first hit is the germline mutation that is inherited in these families(27). A two-hit theory for cancer (Knudsen's two hit hypothesis) for retinoblastoma was first published in 1971, where in the dominantly inherited form of retinoblastoma one mutation is inherited through the germ-cells and a second mutation occurs in somatic cells; in the non-inherited (sporadic) form both mutations

occur in somatic cells(28). It is now clear that many more than two hits are required for a cancer to develop(29) and this is now referred to as the progressive multi-hit model of gene mutation(30); and that initial gene mutations promote genetic instability and therefore promote the occurrence of further abnormalities. This theory also encompasses that the same gene can be involved in a familial cancer as well as in a sporadic form of the disease.

Acquired cytogenetic abnormalities characteristically feature in many types of HMs. These abnormalities can be classified as numerical or structural chromosomal abnormalities. Of the numerical abnormalities, there is a specific pattern as to which chromosomes are lost or gained regardless of the type of HM. Also, for the structural abnormalities that occur, certain genes are frequently involved regardless of the type of HM. It should be noted that people born with constitutional disorders of these genes or chromosomes that are commonly mutated in HMs are also at an increased risk of developing a HM(31). The classic example is trisomy 21; people born with this have a substantially increased risk of developing leukaemia and trisomy 21 is commonly found in HMs.

A known predisposing factor for the development of a HM is a family history of a HM(3-5, 32-34), suggesting a genetic risk for these types of disorders. There are many reports of familial HMs(4-15, 17-24, 33, 34), particularly for one subtype of HM, chronic lymphocytic leukaemia (CLL)(4, 6-8, 11, 15, 18, 22, 24, 34). The relative risk for relatives of people with CLL has been reported to be 7.52, and 1.45 for non-Hodgkin lymphoma (NHL) and 2.35 for Hodgkin lymphoma (HL)(6). An interesting aspect of CLL is that it is the most common leukaemia in western countries(6) but rarely seen in Asian countries(35), and if a person of Asian ethnicity

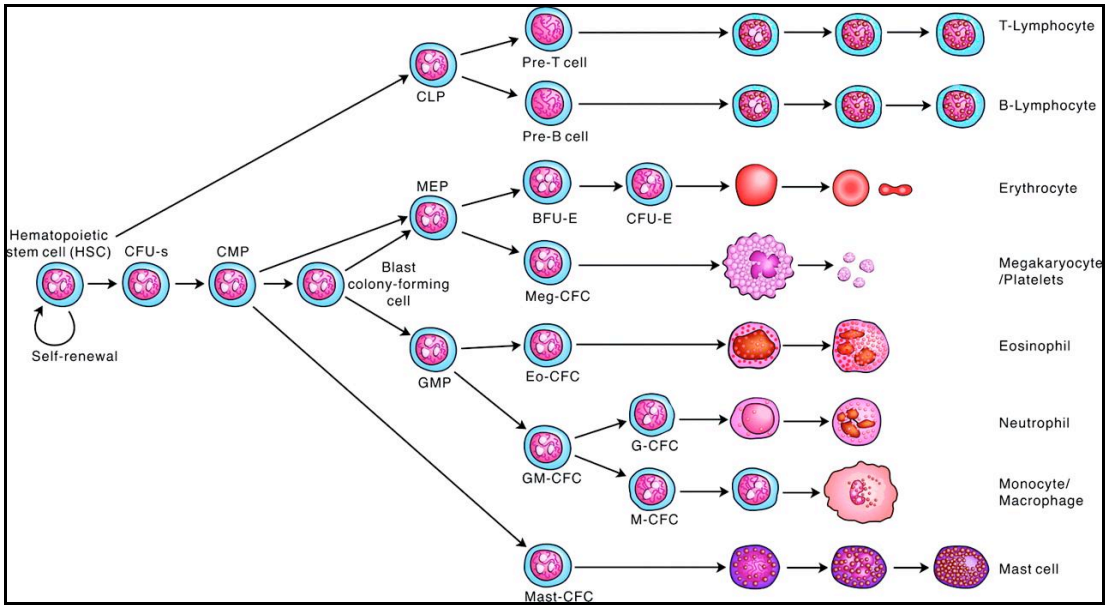
moves to a western country they keep their low risk(35).This supports a genetic rather than an environmental risk with exposure of all family members to the same environmental risk factors(35). There is also a reported increased risk (double) for identical twins if one twin develops a HM by the age of 15(36). This strongly supports a role for genetic factors this disease.

It is hypothesized that in these HM families there is a predisposition allele and if additional acquired mutations occur then a HM develops. The type of HM depends upon which cell receives a second, third or fourth hit(37-39).

1.2. Haematopoietic stem cell development

HSC are derived from cells in the aortic endothelium(40). During embryogenesis production of HSC occurs in the yolk sac, the aorta-gonad mesonephros, the placenta, the fetal liver, the thymus and by the time of birth, from the bone marrow(41). In adult life they are found only in the bone marrow of healthy individuals(41). There are different types of HSC, mostly classified as long term HSC or short term HSC (refer to Figure 1.2.1).

Figure 1.2.1: Development of mature haematopoietic cells from HSC (26).



Multipotential stem cells are self-generating and also produce precursor cells with increasing restriction of their lineage and proliferation potential. (CFU-s Colony forming units, CMP common myeloid progenitors, CLP common lymphoid progenitors, MEP megakaryocyte-erythroid progenitors, GMP Granulocyte macrophage progenitors, BFU blast forming unit, E Erythrocyte, Meg Megakaryocyte, Eo Eosinophil, GM Granulocyte macrophage, M macrophage, G granulocyte).

1.2.1. Types of HSCs

Long term HSC: This is a pluripotent stem cell that is capable of self-renewal as well as maturing to short term HSC. Transcription factors that regulate this type of HSC are *RUNX1*, *SCL/TAL1*, *LMO1*, *MLL*, *TEL(ETV6)*, *BMI-1*, *GFI-1* and *GATA-2*(41). Not surprisingly many of these same genes are involved in many of the leukaemia-associated translocations(41).

Short term HSC: These are referred to as multipotent progenitors, as they can differentiate down any cell line and produce the following common myeloid

progenitor and common lymphoid progenitor. The common myeloid progenitor differentiates into either a megakaryocyte/erythroid progenitor (committed to either the erythroid or megakaryocytic cell lines) or a granulocyte/macrophage progenitor (committed to the myeloid lineage, but can differentiate into a mast cell, eosinophil, neutrophil or a monocyte). The common lymphoid progenitors are committed to the lymphoid cell line, but can still mature into either a T or B lymphocyte.

1.2.2. Reprogramming of committed HSC

It used to be considered that once committed to a specific lineage, the programming of a HSC was unidirectional; this has now been found not to be the case(41). For example, the common lymphoid progenitor can be reprogrammed by expression of *C/EBP* to become a granulocyte/macrophage progenitor and a precursor T cell can be reprogrammed with the expression of *GATA3* to become a mast cell(41). This highlights how related these morphologically distinct haematopoietic cells really are.

1.3. Classification of HMs

The classification of HMs has changed over the years and the current accepted classification system is the WHO classification of HM which was first published in 2001(42) and updated in 2008(1). This classification system takes into account clinical, morphological, immunophenotypic and genetic changes to determine the subtype of HM. The WHO classification system recognises that previous classification systems that relied purely on morphology were insufficient and that some subtypes could not be distinguished by morphology alone(42). Even today the sub-classification of HMs can be difficult and is limited by what tests are available in

a routine diagnostic laboratory. There are 12 broad categories of HMs (see Table 1.3.1) with over 100 subtypes.

Table 1.3.1: Overview of the types of HMs.

Broad category	Number of major subtypes
Myelodysplastic syndromes	7
Myeloid and Lymphoid neoplasms with eosinophilia and abnormalities of *PDGFRA*, *PDGFRB* or *FGFR1*	3
Myelodysplastic/myeloproliferative neoplasms	4
Myeloproliferative Neoplasms	8
Acute myeloid leukaemia	7
Acute leukaemia of ambiguous lineage	7
Precursor lymphoid neoplasms	9
Mature B cell neoplasms	28
Mature T and NK cell neoplasm	17
Hodgkin lymphoma	5
Immunodeficiency-associated lymphoproliferative disorders	4
Histiocytic and dendritic cell neoplasms	7

1.3.1. World Health Organization Classification of HMs

1.3.1.1. Myelodysplastic syndromes

Myelodysplastic syndromes (MDS) are acquired bone marrow failure disorders and are often considered to be "pre-leukaemic" conditions(1) with a risk of transforming to AML. They most commonly present in the 8th decade of life with a predominance of males(1). There are known environmental factors such as benzene exposure and a history of previous chemotherapy or radiotherapy that can cause MDS, however such a predisposing factors account for only 2% of cases of MDS(42). There are 7 types of *de novo* MDS

- Refractory cytopenia with unilineage dysplasia.
- Refractory anaemia with ring sideroblasts.
- Refractory cytopenia with multilineage dysplasia.
- Refractory anaemia with excess blasts.
- MDS with isolated del(5q).
- MDS unclassifiable.
- Childhood MDS.

Classic cytogenetic changes found in MDS include deletions/monosomies of chromosomes 3, 5, 7 and gains of chromosomes 8, 11, 21. All of these abnormalities are also observed in other types of HMs(1).

It has also emerged that epigenetic alterations play a major role in the acquired changes that occur in MDS(43). These changes consist of DNA methylation and histone modification and are capable of changing gene expression(44). A common finding in MDS is hypermethylation of genes involved in cell cycle control and cell adhesion(44). This is of great clinical interest as epigenetic changes are amenable to

reversal with pharmacological drugs, several of which (histone deacetylase inhibitors and DNA methyltransferase inhibitors) are approved for use in MDS(43).

1.3.1.2. Myeloid and Lymphoid neoplasms with eosinophilia and abnormalities of *PDGFRA*, *PDGFRB* or *FGR1*

There are three disorders in this category

- Myeloid and lymphoid neoplasms with *PDGFRA* rearrangements.
- Myeloid and lymphoid neoplasms with *PDGFRB* rearrangements.
- Myeloid and lymphoid neoplasms with *FGR1* rearrangements.

This new category reflects the fact that one genetic mutation that usually results in a myeloproliferative neoplasm can sometimes manifest as a lymphoid neoplasm. The underlying factor in this category is a genetic defect that results in aberrant tyrosine kinase activity. It highlights that one genetic mutation that can result in different types of clinical diseases.

1.3.1.3. Myelodysplastic/Myeloproliferative disorders

This was a new category to this classification syndrome in 2001. The inclusion of this category highlights the clinical heterogeneity of HMs, in that they can represent a spectrum of disease features rather than a specific disease(1). There are 4 subtypes in this category.

- Chronic myelomonocytic leukaemia.
- Atypical chronic myeloid leukaemia BCR-ABL1 negative.
- Juvenile myelomonocytic leukaemia.
- MDS/MPD unclassifiable.

Chronic myelomonocytic leukaemia has features of both MDS and MPD with the degree of dysplasia/proliferation different in each patient. However the common acquired cytogenetic findings are not unique to this disorder and overlap with MDS, MPD and AML abnormalities(1).

1.3.1.4. Myeloproliferative Neoplasms (MPD)

According to the WHO classification syndrome there are 8 MPDs

- Chronic myeloid leukaemia.
- Chronic neutrophilic leukaemia.
- Chronic eosinophilic leukaemia NOS.
- Essential thrombocytosis.
- Polycythaemia vera.
- Myelofibrosis.
- Mastocytosis.
- Myeloproliferative neoplasm unclassifiable.

These disorders are associated with an increase (proliferation) in circulating haematological cells: platelets, granulocytes or red cells(1). However many patients have more than one cell type increased. Intriguingly, it was postulated in 1951 that a group of disorders (chronic granulocytic leukemia, polycythemia vera, idiopathic myeloid metaplasia of the spleen, thrombocythemia, megakaryocytic leukemia and erythroleukemia) should be classified together because of their overlapping clinical features (45). These disorders still form the basis of this category and this concept still appears valid even with today's knowledge.

Chronic myeloid leukaemia (CML) is defined by a specific translocation (t(9;22)(q34;q11.2)), which results in the fusion of the *BCR* and *ABL1* genes and produces an abnormal tyrosine kinase with deregulated activity(46). This disorder is associated in the literature with a series of firsts; it was the first type of leukaemia to recognised as a distinct clinical entity (46), the first HM to be defined by its cytogenetic cause(47, 48), the first HM to have the genes identified and the first HM to be treated with a targeted therapy against the specific acquired genetic mutation(49). More recently, mutations in another tyrosine kinase receptor (*JAK2*) have been found in MPDs(50). This mutation in *JAK2* (V617F) is not specific to a particular type of MPD, but can be found in many. It has been reported that the type of HM that develops is dependant upon the allelic burden, however this dosage effect does not explain fully the phenotypic heterogeneity of MPDs(51). This once again highlights how the same genetic mutation is associated with different types of HMs.

1.3.1.5. Acute myeloid leukaemia and related precursor neoplasms

There are approximately 7 major subtypes of AML

- AML with recurrent genetic abnormalities.
- AML with myelodysplasia-related changes.
- Therapy related AML.
- AML NOS.
- Myeloid sarcoma.
- Myeloid proliferations related to Down syndrome.
- Blastic plasmacytoid dendritic cell neoplasm.

AML used to be defined as having a blast cell count greater than 30% in the bone marrow, however with the new classification system(1) the minimum blast count is reduced to 20%. In a few specific cytogenetic subtypes [t(8;21)(q22;q22), inv(16)(p13.1q22) or t(16;16)(p13.1;q22), t(15;17)(q22;q12), t(9;11)(p22;q23), t(6;9)(p23;q34), inv(3)(q21q26.2), t(3;3)(q21;q26.2) and t(1;22)(p13;q13)] the blast cell count is irrelevant and in these types it is classified as AML, whereas previously they would have been classified as a MDS. Under the subtype of AML with recurrent genetic abnormalities, a new subtype emerged in the current classification, AML with gene mutations (fms-related tyrosine kinase 3; nucleophosmin; *CEBPA*; *KIT*; *MLL*; *WT1*; *NRAS*; *KRAS*). Mutations in these genes can occur in any subtype of AML(1). There have been several reports of familial AML(52) in the literature, mostly with mutations involving *RUNX1*(53) or *CEBPA*(21, 52), the same genes that are involved in mutations in non-familial AML. Many of the classic secondary changes (such as del3q, del5q, del7q, monosomy 7, trisomy 8) that are seen in AML are seen in other types of HM(1), also many of the genes (*MLL*, *RUNX1*) translocated in AML are also translocated in other types of HMs.

1.3.1.6. Acute leukaemia of ambiguous lineage

There are 7 major groups of acute leukaemia with ambiguous lineage(1)

- Acute undifferentiated leukaemia.
- Mixed phenotype acute leukaemia with *BCR-ABL1*.
- Mixed phenotype acute leukaemia with *MLL* rearrangement.
- Mixed phenotype acute leukaemia B/Myeloid NOS.
- Mixed phenotype acute leukaemia T/Myeloid NOS.
- Mixed phenotype acute leukaemia NOS.

- Natural killer cell lymphoblastic leukaemia/lymphoma.

This category of leukaemia again highlights the overlap that can occur between the different cell lineages and it occurs because the leukaemogenic event occurred in a primitive HSC that still had the ability to either develop into a granulocyte, B lymphocyte, T lymphocyte or NK cell. This again shows that there is an overlap in the genetic events that can cause different HMs.

1.3.1.7. Precursor lymphoid neoplasms

There are two major categories of precursor lymphoid neoplasms, either B lymphoblastic leukaemia/lymphoma or T lymphoblastic leukaemia/lymphoma(1). The B lymphoblastic leukaemia/lymphomas are further classified according to the genetic abnormalities found. Their names show the arbitrary divide between lymphoma and leukaemia, in that the same disease is called one or the other depending on the predominant site of disease, either in the lymph nodes or solid tissue (lymphoma) or bone marrow or blood (leukaemia).

B lymphoblastic leukaemia/lymphoma

These disorders are seen in both adults and children, with a peak incidence in children between the age of 2-5 years(1). There are 8 types of B lymphoblastic leukaemia/lymphoma mostly based on genetic changes

- B lymphoblastic leukaemia/lymphoma NOS.
- B lymphoblastic leukaemia/lymphoma with t(9;22)(q34;q11.2).
- B lymphoblastic leukaemia/lymphoma with *MLL* rearrangement.
- B lymphoblastic leukaemia/lymphoma with t(12;21)(p13;q22).

- B lymphoblastic leukaemia/lymphoma with hyperdiploidy.
- B lymphoblastic leukaemia/lymphoma with hypodiploidy.
- B lymphoblastic leukaemia/lymphoma with t(5;14)(q31;q32).
- B lymphoblastic leukaemia/lymphoma with t(1;19)(q23;p13.3).

These genetic subtypes each have a different prognosis and treatment options, but morphologically the lymphoblasts are indistinguishable. They also overlap with other HMs in that the genes *(BCR/ABL1, MLL, RUNX1*) that are involved in defining subtypes of ALL are also commonly involved in causing other types of HMs.

T lymphoblastic leukaemia/lymphoma

This disorder is common in adolescent males and often presents with a mediastinal mass(1). Approximately 50-70% of cases will have an abnormal karyotype and they typically involve recurrent rearrangements of the T cell receptor loci (14q11.2 and 7q35)(1). The distinction between lymphoma and leukaemia in this disorder is arbitrary; the term 'lymphoma' if there is purely nodal disease and 'leukaemia' if blood or bone marrow is predominantly involved. This highlights how cells of the same phenotype can give rise to differing clinical manifestations. This disorder generally has a poorer prognosis than B-ALL and as yet is not classified based upon the specific genetic subtypes.

1.3.1.8. Mature B cell neoplasms

There are 28 major types of mature B cell neoplasms according to the WHO(1)

- Chronic lymphocytic leukaemia/small lymphocytic lymphoma
- B-cell prolymphocytic leukaemia.
- Splenic marginal zone lymphoma.

- Hairy cell leukaemia.
- Splenic lymphoma/leukaemia unclassifiable.
- Lymphoplasmacytic lymphoma.
- Heavy chain disease.
- Plasma cell neoplasms.
- Extranodal marginal zone lymphoma of mucosa-associated tissue.
- Nodal marginal zone lymphoma.
- Follicular lymphoma.
- Primary cutaneous follicle centre lymphoma.
- Mantle cell lymphoma.
- Diffuse large B cell lymphoma (DLBCL), NOS.
- T cell/histiocyte rich large B-cell lymphoma.
- Primary DLBCL of the CNS.
- EBV positive DLBCL of the elderly.
- DLBCL associated with chronic inflammation.
- Lymphomatoid granulomatosis.
- Primary mediastinal large B cell lymphoma.
- Intravascular large B cell lymphoma.
- ALK positive large B cell lymphoma.
- Plasmablastic lymphoma.
- Large B cell lymphoma arising in HHV8-associated multicentric Castleman disease.
- Primary effusion lymphoma.
- Burkitt lymphoma.

- B cell lymphoma unclassifiable with features intermediate between DLBCL and Burkitt lymphoma.
- B cell lymphoma unclassifiable with features intermediate between DLBCL and classical Hodgkin lymphoma.

They consist of diseases that affect mature B cells and group together what used to be called the lymphoproliferative disorders (chronic leukaemia) and the non-Hodgkin lymphomas (NHL). In fact the divide between these phenotypic groups is non-existent, in that CLL can also be called small lymphocytic lymphoma (SLL) depending upon where the bulk of the disease presents (either in the lymph nodes (SLL) or circulating in the blood (CLL) or even from where the diagnostic material is taken: in other words a lymph node biopsy from a CLL patient will be reported as showing SLL if the pathologist is unaware of the blood findings. This once again highlights the overlap that exists between the different types of HMs.

The majority of mature B cell neoplasms typically present in the 7th decade of life with a male predominance although they are found in all ages(1). Several of the major subtypes are defined by specific chromosomal translocations: Mantle cell lymphoma by the t(11;14)(q13;q32) involving the *CCND1* and *IGH@* genes however this translocation is also seen in multiple myeloma (MM) and plasma cell leukaemia(54); Follicular lymphoma by the t(14;18)(q32;q21) involving the *IGH@* and *BCL2* genes which can also be seen in DLBCL and Burkitt lymphoma(55, 56); Burkitt lymphoma by an 8q24 (*MYC*) rearrangement which can also be seen in DLBCL(57), CLL and MM. This shows how one genetic abnormality can give rise to many different types of HMs, which supports the hypothesis that a germline

mutation can give rise to different types of HMs depending upon which HSC the next hit arises in.

1.3.1.9. Mature T cell neoplasms

This group of HMs links disorders that used to be separated as T cell lymphomas and T cell lymphoproliferative disorders(1). Mature T cell neoplasms are less frequent than mature B cell neoplasms, but in the WHO classification there are 17 major types of mature T cell neoplasms(1)

- T-cell prolymphocytic leukaemia.
- T-cell large granular lymphocytic leukaemia.
- Chronic lymphoproliferative disorders of NK cells.
- Aggressive NK cell leukaemia.
- Epstein –Barr virus positive T cell LPD of childhood.
- Adult T-cell leukaemia/lymphoma.
- Extranodal NK/T cell lymphoma nasal type.
- Enteropathy associated T cell lymphoma.
- Hepatosplenic panniculitits like T cell lymphoma.
- Mycosis fungoides.
- Sezary syndrome.
- Primary cutaneous CD30 positive T cell LPDs.
- Primary cutaneous peripheral T cell lymphoma rare subtypes.
- Peripheral T cell lymphoma NOS.
- Angioimmunoblastic T cell lymphoma.
- Anaplastic large cell lymphoma ALK positive.
- Anaplastic large cell lymphoma ALK negative.

1.3.1.10. Hodgkin lymphoma

Hodgkin lymphoma was first described in 1832, but only in the last 10 years have the malignant cells (Hodgkin and Reed–Sternberg cells) been recognised as B lymphocytes(1). There are two major types of HL

- Nodular lymphocyte predominant HL.
- Classical HL (CHL).

CHL is subdivided into 4 types: Nodular sclerosis CHL, Mixed cellularity CHL, Lymphocyte-rich CHL and Lymphocyte depleted CHL.

1.3.1.11. Others

Both of the two other major subtypes of HMs, immunodeficiency associated LPDs and histiocytic and dendritic cell neoplasms are very rare(1).

1.4. Recurrent acquired cytogenetic abnormalities observed in more than one type of HM.

The acquired cytogenetic changes that occur in HMs have been widely studied for many years, and in fact most of the reported cytogenetic changes associated with cancers were first reported in HMs(58). Despite the fact that there are over 100 different subtypes of HMs(1), many of the acquired genetic changes that occur are not unique to one subtype of HM, but are seen in many different subtypes. This is true for both specific genes and classic chromosome abnormalities.

1.4.1. Specific genes

1.4.1.1. *RUNX1* translocations

The *RUNX1* gene is located on the long arm of chromosome 21 at region 2 band 2 (chr21q22). This gene was originally called *AML1* due to its translocation with *RUNX1T1* [t(8;21)(q22;q22)] in the leukaemic cells of patients with AML. This translocation results in a specific type of AML (AML with maturation), now classified as AML with t(8;21)(q22;q22)(1). However this gene (*RUNX1*) is also the most commonly translocated gene in childhood ALL [t(12;21)(p13;q22)](1). The normal function of *RUNX1* involves it forming a heterodimer with core-binding factor beta. A rearrangement of core-binding factor beta is associated with a specific subtype of AML, AML with eosinophilia with inversion (16)(p13.1q22)(1). Acquired point mutations in *RUNX1* have also been found in AML and ALL(1).

1.4.1.2. *BCR/ABL1* fusions

The translocation that results in the fusion of the *BCR* gene on the long arm of chromosome 22 with the *ABL1* gene on the long arm of chromosome 9 defines CML(59). It is interesting to note that this translocation (t(9;22)(q34;q11.2)) is also seen in 5% of children with ALL, 15-30% of adults with ALL and 2% of patients with de novo AML(60). Even when the molecular breakpoints of the translocation are analysed, they are not specific for CML(60). There are 3 main breakpoints in the *BCR* gene, namely minor *BCR,* major *BCR* and micro *BCR*. Depending upon where the breakpoint is, a different sized fusion protein is transcribed. The minor *BCR* (a1a2 breakpoints) results in a $p^{190BCR\text{-}ABL}$, whereas the major *BCR* (either b2a2 or b3a2) results in a $p^{210BCR\text{-}ABL}$, and the micro *BCR* (e19a2) results in a $p^{230BCR\text{-}ABL}$(60).

These discoveries show that the same genetic mutation can give rise to different types of HMs.

1.4.1.3. Mixed Lineage Leukaemia *(MLL)* translocations

Translocations involving this gene (*MLL*) are seen in both myeloid and lymphoid disorders (61) and has over 60 partner genes that have been reported in HMs(62). However each different translocation partner is not specific for a particular type of HM(61), and the same translocation can be seen in different types of HMs(61). The classic example would be the translocation between *MLL* and *AF4* [t(4;11)(q21;q23)], which is seen in infant acute ALL, but can also be seen in AML(63).

1.4.1.4. *IGH@* translocations

The *IGH@* locus is rearranged in most B lymphoid malignancies and is involved with many other partner genes(1). However a few specific *IGH@* translocations are regarded as hallmarks of a particular subtype of HM, in particular the *IGH@/CCND1* translocation that defines MCL although this translocation can also be seen in MM(54). Similarly, the *IGH@/MYC* translocation that defines Burkitt lymphoma is not specific to Burkitt lymphoma but can also be seen in other types of lymphoproliferative disorders(56). The *IGH@/BCL2* translocation that defines follicular lymphoma is also not specific and can also be seen in DLBCL(56).

1.4.2. Common chromosomal abnormalities

Many of the common chromosomal changes that are seen in HMs are also not specific to one subtype of HM.

1.4.2.1. Trisomy 8

Trisomy 8 is observed in both myeloid and lymphoid disorders and both chronic and acute types of HMs and is one of the most common secondary chromosomal gains seen in HMs(1). It should also be noted that patients with Warkany syndrome, which is defined as a mosaic trisomy 8 constitutional disorder, are at increased risk of developing a HM(64). It has also been reported that up to 15-20% of cases of HMs that have trisomy 8 as the sole abnormality, are due to an underlying Warkany Syndrome rather than an acquired gain(64).

1.4.2.2. Sex chromosome abnormalities

Sex chromosome abnormalities (monosomies or trisomies) are very common secondary changes and are found in both myeloid and lymphoid malignancies(65).

1.4.2.3. Trisomy 21

Trisomy 21 as an acquired gain is frequently seen in ALL(66) and is the second most common trisomy in myeloid disorders(67).

1.5. Constitutional disorders predisposing to develop a haematological malignancy

Constitutional cytogenetic disorders are generally classified as an aneuploidy (monosomy or trisomy) or structural disorders (translocations, deletions, duplications or inversions) of chromosomes that a person is born with. Many of these disorders also have an increased risk of developing a HM(68-70). It seems that the underlying genetic abnormality acts as the first hit and provides an environment that makes the cells prone to acquiring more genetic abnormalities.

1.5.1. Aneuploidy disorders

1.5.1.1. Trisomy 21

This is the most common autosomal trisomy and is observed in 1 in every 600-1000 live-born births(71). Children born with this disorder are at a 100 times increased risk of developing a HM, in particular acute megakaryoblastic leukaemia, which is otherwise an extremely rare subtype of acute myeloid leukaemia(72). They can also develop transient abnormal myelopoiesis (TAM), which morphologically looks like AML, but has a high rate of spontaneous remission(73). Approximately 20% of these TAM patients will then go on to develop acute megakaryoblastic leukaemia by the age of 4 years(69) with acquired *GATA1* mutations. In the current classification system, the myeloid proliferations related to Down syndrome are grouped together in their own category(1).

1.5.1.2. Other autosomal trisomies

Only a few other autosomal trisomies are compatible with life. These are trisomy 18, trisomy 15, trisomy 13, trisomy 22 and mosaic trisomy 8(71). Patients with these disorders are also at an increased risk of developing a HM(68). Furthermore these trisomies may also be seen as secondary gains in many different types of HMs(67). Trisomy 8 is seen in approximately 1 in 25000 livebirths(69), but is the most common trisomy seen in HMs. Interestingly a study found that in approximately 15-20% of cases, a sole trisomy 8 in a HM is actually a constitutional finding rather than an acquired gain(74). Trisomy 22 is the most common secondary finding in AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22)(75).

1.5.1.3. Numerical sex chromosome abnormalities

Turner syndrome

Turner syndrome is seen in approximately 1 in 2000 livebirths(76) and is defined as 45,X. A non-significant increase in leukaemia is seen in people with Turner syndrome(76). A loss of an X chromosome as a secondary cytogenetic event is seen as a frequent event in HMs(77).

Other disorders

Loss of the Y chromosome is also a common acquired finding in HMs(78). It is unclear whether this loss is purely age related, as the incidence of –Y increases with age, or specifically disease associated(78). If over 75% of cells have a –Y, then it is thought to be disease related(79).

1.5.2. Structural chromosomal abnormalities

Beckwith-Weidemann syndrome (11p11-15.1)

Beckwith-Weidemann Syndrome (BWS) is a fetal overgrowth syndrome with an increased incidence of childhood cancers, in particular Wilms tumours(80). The genetics of BWS are complex and can arise by duplication, uniparental disomy (UPD), inversion or translocation of chromosome 11p11-15, however what is common is that BWS arises if there is no maternal contribution to 11p15(80). Patients with this disorder have been shown to have a significant excess of cancer deaths(81). In Wilms tumours it is noted that there is loss of the maternal chromosome 11 in the tumour cells(80). Acquired UPD for many regions of the genome is now being found in many types of HMs(82).

Prader-Willi/Angelmen Syndrome

Both of these disorders can be caused by the same genetic deletion of chromosome 15q11-13(83), but they have completely different disease phenotypes. This shows that even in constitutional disorders the same genetic defect can result in different disorders. Prader-Willi syndrome (PWS) is caused by a deletion of chromosome 15q11-13 in approximately 60% of cases(83) and is seen in approximately 1 in 1600 births(84). The phenotype consists of a floppy baby, developmental delay, intellectual impairment and severe childhood obesity (80). A deletion of chromosome 15q11-13 is also seen in 60% of cases of Angelmen syndrome(83). Patients with Angelmen syndrome are often described as a "happy puppet", due to their happy nature, jerky limb movements and severe mental impairment(83). The chromosome deletion in these two syndromes can be at the exact same breakpoints, but if the deletion occurs on the chromosome 15 inherited from the father then the patient has PWS, whereas if the deletion is in the chromosome 15 inherited from the mother then it results in Angelmen syndrome, this is due to differential expression of imprinted genes which have different maternal or paternal methylation patterns(85). Acquired changes in methylation is seen as a common event in HMs(86). Changes in methylation of *DAPK1* resulting in down regulation of *DAPK1* expression have been implicated in both sporadic and familial CLL(25). Thus the same mechanism (changes in methylation) can result in different types of disorders, both constitutional and malignant(87) and the same chromosomal genetic disorder can result in completely different diseases (phenotypes).

Miller Dieker Syndrome

This is a contiguous gene deletion syndrome that is associated with lissencephaly, severe mental retardation and deletions of chromosome 17p13.3(88). Deletions of chromosome 17p are one of the commonest abnormalities seen in all types of

cancers(89) and *TP53* (located at 17p13.1) is sometimes, but not always implicated as the gene involved(90). Two other tumour suppressor genes are in this region (*HIC1* and *OVCA1*) and case reports of cancers occurring in patients with Miller Dieker syndrome are emerging in the literature(90). It is now recommended that surveillance for malignancies be part of the long term management of these patients(91). Such observations highlight the fact that structural genetic disorders can result in an increased susceptibility to developing a malignancy. A case report of a girl with Miller Dieker syndrome developing ALL has been reported in the literature(92) and has recommended that surveillance for malignancies be a part of management for these patients.

1.6. Specific single-gene disorders with a known increased risk of a HM

1.6.1. *TP53* Li-Fraumeni syndrome

This syndrome was first described by both Li and Fraumeni in 1969(93). It is now defined as a proband aged under 45 years with any childhood tumour, or a sarcoma, brain tumour, or adrenocortical tumour plus a first or second degree relative in the same lineage with a typical Li-Fraumeni tumour at any age and an additional first or second degree relative in the same lineage with any cancer under the age of 60(94). Over 200 different germline mutations in this gene have been described(94), however the five most common germline mutations in *TP53* are also the most common mutations found in sporadic cases of these tumours(94). *TP53* mutations are frequently seen in HM and are associated with a poorer outcome(95).

1.6.2. Dyskeratosis congenita

This disease can be caused by several different mutations. The X-linked recessive form of this disease has been shown to be due to mutations in *DKC1*(96). This gene encodes dyskerin, which is a key component in stabilising the telomerase complex(96). In contrast, the autosomal dominant form of this disease has been shown to be due to mutations in *TERC* or *TERT*(96). Both of the genetic causes (X-linked or autosomal) result in a bone marrow failure syndrome and increased malignancy risk(96).

1.7. Familial Haematological disorders

1.7.1. Fanconi anaemia

This disorder was first described in 1927 by Guido Fanconi(97) as a syndrome of bone marrow failure, short stature and radial bone abnormalities. This disease can be either autosomal or X linked and mutations have been described in at least thirteen different genes (*FANCA, FANCB, FANCC, BRCA2, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCJ, FANCL, FANCM, PALB2*)(98). The genes involved in causing Fanconi anaemia are associated with cell cycle checkpoint control or DNA repair pathways(99). Patients with Fanconi anaemia are at a high risk of both HMs and solid malignancies(98).

1.7.2. Familial platelet disorder

This is an autosomal dominant familial platelet disorder characterised by thrombocytopenia and a tendency to develop AML(53), which has been shown to be due to point mutations in *RUNX1*(53). For the AML to develop, a second mutation

has to occur, either in *RUNX1* or another gene(53). This supports the hypothesis that a germline mutation in these families is the first hit in a series of hits that must occur for a HM to develop. Other types of leukaemia (lymphoid) have also been reported in carriers of the *RUNX1* mutations(100).

1.7.3. Congenital Neutropenia

This is a heterogeneous condition characterised by bone marrow arrest of myelopoiesis at the promyelocyte/myelocyte stage with an absolute neutrophil count of $<0.5x10^9$/L(101). One particular type of congenital neutropenia is Kostmann syndrome, which was first described by Rolf Kostmann as a recessive lethal disease; prior to this it was thought to be an acquired disorder(102). It has been found that approximately 80% of patients with congenital neutropenia develop acquired mutations in the *G-CSFR* gene(101) and go on to develop AML, once again supporting the hypothesis that germline mutations are the first hit in a series of hits required to develop a HM and that disorders that were thought to be sporadic can be due to germline mutations.

1.8. Known causes of malignancy

It is known that DNA damage can cause cancer (103). Cells are under constant attack from DNA damaging agents such as UV radiation; both DNA repair pathways and cell cycle checkpoint pathways have evolved to limit this damage(103). How a cell responds to this damage is critical to the risk of a person developing a malignancy. It is known that multiple genetic changes are needed for a cell to undergo a cancerous change(104). Genetic changes occur in genes that control how cells progress through the cell cycle or apoptosis pathways(104) and these pathways are

interconnected(105). Classically genes involved in causing cancer have been classified as either tumour suppressor genes or oncogenes; however it is now known that individual genes can in fact be both(106). Furthermore it is now emerging from whole genome sequencing that mutations are present in genes that were previously not thought to be involved in the pathogenesis of cancer(107).

1.9. Use of families to find genetic causes of diseases

Genetic studies using families have been successful in identifying the causative genetic mutation in many diseases and familial cancer syndromes(108). It has been found that the many genes identified in causing familial cancers are also the same genes mutated in sporadic forms of the same cancers. Thus the study of familial cancers can led to insight into disease development in the general population. It establishes a justification for using families as a means to finding genetic mutations(109). Familial studies compare the genetic makeup of affected individuals to see what regions they share (linkage). This was originally done using microsatellites, but now is being done with single nucleotide polymorphisms (SNPs). This identifies regions of the genome that can then be sequenced for mutations.

1.9.1. Genetic mutations discovered using Tasmanian families

Tasmanian families have been used in many genetic studies(33, 110-115). Tasmania is a unique resource for genetic studies as it is an island state with a relatively stable population, 65% being descendants of colonial settlement, a centralized cancer registry and centralized specialised pathology testing.

Tasmanian families were used in genetic linkage studies investigating the genetic basis of multiple endocrine neoplasia type 1 (MEN1) which is an autosomal

dominant disorder(116). This lead to the identification of a gene on chromosome 11(117). This disorder is another example of a germ line mutation in a gene (*MEN1*) giving rise to multiple different tumours of the endocrine tumours and tumours of the skin and central nervous system (gastrinomas, insulinoma, glucagonoma, pituitary ademoma, adrenocortical carcinoma, angiofibromas, collagenomas, lentiginosis, rhabdomyosarcoma, leiomyosarcoma, renal carcinomas and thyroid carcinomas)(108). Studies on Tasmanian families with prostate cancer have led to the identification of a locus on chromosome 5(111).

Tasmanian families with increased incidence of non-malignant medical diseases have been used to identify the underling genetic causes of their disease, with a locus involved in glaucoma, identified on chromosome 10q22(110).

1.9.2. Familial cancer syndromes

The concept that cancer could be inherited has been postulated for over 100 years(118). The identification of the specific mutations that cause familial cancer has had a direct benefit on the affected families in enabling genetic testing and counselling. The most common familial cancers that have diagnostic testing available include breast cancer, bowel cancer, ovarian cancer and thyroid cancer(119). There are now approximately 100 genes that have been identified that cause Mendelian inherited cancer syndromes(118). It is only with the knowledge of the specific genetic mutations that appropriate surveillance, counselling and treatment can be implemented for these families.

1.9.2.1. Breast Cancer

Linkage studies of families with early onset breast cancer lead to the identification of a disease locus, located on chromosome 17q21(120), subsequently sequencing of this region found mutations in the breast cancer 1, early onset gene (*BRCA1)*(120). Subsequently the *BRCA2* gene was identified on chromosome 13(121). These discoveries have changed the management of BRCA-positive familial breast cancer, including more intensive screening in those possessing the mutated genes. From these studies three classes of predisposition risk for breast cancer have emerged: firstly people with *BRCA1* and *BRCA2* mutations have high-penetrance predisposition alleles; secondly people with mutations in *CHEK2*, *ATM*, *BRIP1* or *PALB2,* which are genes functionally related to *BRAC1* and *BRAC2,* who have an intermediate risk; and thirdly genome-wide association studies have identified 8 variants that have a low penetrance risk of breast cancer(122). Mutations in *BRCA1* and *BRCA2* are reported to account for 20% of the heritable risk; mutations in *CHEK2*, *ATM*, *BRIP1* or *PALB2* account for approximately 5% of the heritable risk(123).

1.9.2.2. Colorectal Cancer

Approximately 30% of patients with colorectal cancer have a familial component to their disease(124). Single gene germline mutations have been have been found in approximately 5-6% of cases of colorectal cancer(125).

Familial adenomatous polyposis

This is an autosomal dominant disease with a 100% risk of developing colorectal cancer by the age of 39 years and is due to mutations in the *APC* gene on

chromosome 5q21(125). Diagnostic testing for mutations in *APC* has been available since 1994 and knowledge of the carrier status with annual colonoscopy/endoscopy and eventually prophylactic total colectomy can dramatically reduce the risk of developing colorectal cancer(125).

Hereditary non-polyposis colon cancer (Lynch syndrome)

This syndrome was first described by Dr Lynch in 1966(126) in two large kindred's which had an aggregation of colorectal and endometrial cancers that were inherited as an autosomal dominant disorder. Lynch syndrome accounts for approximately 3-5% of colorectal cancers and is caused by mutations in one of four DNA mismatch repair genes (*MSH2*, *MLH1*, *MSH6* or *PMS2*)(125). Knowledge and testing for carrier status has dramatically improved survival in these patients.

1.9.2.3. Von Hippel-Lindau syndrome

Von Hippel-Lindau syndrome results from mutations in the *VHL* tumour suppressor gene and shows marked clinical variation with age dependent penetrance(127). The most common clinical manifestations are retinal and central nervous system hemangioblastomas, renal cell carcinomas, pheochromocytoma, pancreatic endocrine tumours and endolymphatic sac tumours(128). Recognition of this syndrome and the understanding of the underlying heritable risk has improved patient outcomes by implementation of screening programs in carriers and genetic testing of relatives(129).

1.10. Familial haematological studies in the literature

In the literature there are 8 major epidemiological studies on the incidence of HM in families that have been reported(130). Three were cohort studies and 5 were case-

controls. The familial relative risk calculated in these 8 studies ranged from 2.4 to 8.5(6, 32, 33, 131-135). One of these studies was conducted in Tasmania(33, 112, 113).

1.10.1. The 1970s Tasmanian cohort study

In the 1970s a cohort of all patients that were diagnosed with a HM in Tasmania was surveyed with publications in the 1970s and two publications in the 1980s (33, 112, 113). At the time it was appreciated that a number of the people affected with a HM were related. These initial studies are the basis for the current families in this study. These studies were focused on documenting the residential and occupational history of the patients; however it became apparent that families with two or more closely related individuals with a MPD or LPD were not uncommon(112) and thus from 1975 formal questioning about relatives was undertaken, and confirmation of disease status and subtype was obtained from clinical colleagues around Australia. A total of 866 people in Tasmania were diagnosed with either a myeloproliferative or a lymphoproliferative disorder between 1 of January 1972 to the 31st of December 1980 and 100% were enrolled in these studies(33, 113). Adult controls were randomly selected from the electoral roll (The Australian Electoral Commission maintains and updates the Commonwealth electoral roll - a list of eligible persons who are registered to vote in Australian election) and matched for age and gender. Paediatric controls were randomly selected from the records of the Child and School Health Services of the Department of Health.

Giles *et al.* 1984 reported that the overall lymphoproliferative risk for relatives of male patients was 3.61 compared to the risk for relatives of female patients of 3.15. The myeloproliferative risk was 5.17 and 5.21 respectively(33). This is a similar

relative risk when compared to other studies of the relative risk of HM. Goldin *et al.* 2006, found that the relative risk for relatives of people with CLL was 7.52(6) but it has been reported by Pottern *et al.* 1991(135), to be as low as 2.3.

1.10.2. Familial studies of CLL

A review article by Yuille *et al.* in 2000 reported that there were 81 families with CLL in the literature(24). It was also noted that in five of these families other family members were affected with other subtypes HMs(24). Mauro *et al.* 2006, reported that approximately 12.5% of cases of CLL have a positive family history of CLL(15). There have been several recent genome-wide linkage studies on familial CLL(18, 136-138) to try and identify the underlying genetic predisposition in this disease.

Sellick *et al.* 2007, reported the linkage analysis from an international consortium that enrolled 206 familial CLL families, with maximum nonparametric linkage found on chromosome 2q21.2 and a further two chromosomal regions (6p22.1 and 18q21.1) which had significant LOD scores(18). The authors of this report, Sellick *et al.* 2007, attest that these findings provide direct evidence for Mendelian predisposition to CLL and that more than one gene is responsible for the genetic risk. Fuller *et al.* 2008, found maximum nonparametric linkage to 2q37, 4q35, 11p15 and 14q24. Ng *et al.* 2007, identified a region on 13q21.33-q22.2 as harbouring a CLL predisposition gene(137). Of all these families with CLL, only one mutation in one family has been published to date: this was a mutation in an enhancer region of *DAPK1*. This mutation resulted in decreased expression of *DAPK1*(25). Decreased expression of *DAPK1* is common in sporadic and familial CLL, but the mutation has

not been found in other families with familial CLL or as an acquired change in sporadic CLL(25).

1.10.3. Anticipation in familial HM.

In a number of HM families, the phenomenon of 'anticipation' has been described(5, 7, 10, 20, 22, 23, 139). Anticipation is defined as the onset of the disease in question at an earlier age or increased disease severity in successive generations, and has implications for the nature of the genetic condition that underpins it.

Anticipation in CLL families has been reported with a reduction in the age at diagnosis by 15 years(7). It has also been observed in AML families, with the mean age at diagnosis in the first generation being 57 years, second generation 32 years and third generation 13 years(10). Anticipation has also been observed in lymphoma families(17), families with plasma cell dyscrasia(5) and for Hodgkin lymphoma(20) with a total of 140 families with HMs being documented as showing anticipation(17), although this phenomenon is not universally observed(140). It is often argued that this phenomenon is due to selection bias, however if this was the case, then it would be present in all families.

Anticipation is reported to be consistent with a dominantly acting mutation**;** however causative mutations in HMs have not been identified. In other disorders where anticipation has been reported, it has been found to be due to an unstable dynamic mutation that expands with each generation(141). This type of mutation has been implicated in over 40 diseases. Triplet repeats have been studied in familial CLL, however no causative repeat was identified(142).

1.10.4. Genome-wide association studies (GWAS)

The number of reported GWAS has increased dramatically over the last few years with the advances in genotyping technology. These studies investigate genomic variation across the entire genome in order to identify genetic alleles that are associated with a measurable trait/disease. These types of studies have become the standard method in recent years to identify genetic causes of common disease(143). GWAS have found that common polymorphisms do contribute to common diseases(143). Several GWAS have been undertaken in the field of HMs(144-148).

1.10.4.1. Genome-wide association studies in CLL

A recent GWAS by Di Bernardo *et al.* 2008, identified six susceptibility loci in CLL(149). These loci were identified using 299,983 tagging SNPs in 1529 cases and 3115 controls(149). The loci identified are 2q13 (rs17483466; $P = 2.36 \times 10^{-10}$), 2q37.1 (rs13397985, *SP140*; $P = 5.40 \times 10^{-10}$), 6p25.3 (rs872071, *IRF4*; $P = 1.91 \times 10^{-20}$), 11q24.1 (rs735665; $P = 3.78 \times 10^{-12}$), 15q23 (rs7176508; $P = 4.54 \times 10^{-12}$) and 19q13.32 (rs11083846, *PRKD2*; $P = 3.96 \times 10^{-9}$)(149). These results suggest that more than one gene is contributing to the risk of developing CLL, but the authors state that these loci only account for approximately 3% of the excess familial risk of CLL(149). A study by Rudd *et al.* 2006, investigated SNPs in the *ATM-BRCA2-CHEK2* pathway and identified 49 SNPs that are significantly associated with an increased risk of developing CLL(150).

Slager *et al.* 2010, reported a replication study of seven SNPs previously identified in a GWAS study and validated them in a further 438 CLL cases and 328 controls(147). Six of those previously identified SNP continued to show a significant association,

however they concluded that further study is needed to identify the causative mutation and to determine how these SNPs contribute to the pathogenesis of CLL(147).

Crowther-Swanepoel *et al.* 2010, undertook a GWAS with validation in four datasets in a total of 2503 CLL cases and 2503 controls(144). They identified four new susceptibility loci for CLL at 2q37.3, 8q24.21, 15q21.3 and 16q24.1(144) and for two previously reported loci.. Therefore, in total 10 loci for CLL have been reported for CLL; however these are reported to account for only 10% of the heritable risk, with many more additional risk variants yet to be discovered(144). These GWAS in CLL continue to provide evidence of multiple genetic predisposition alleles for familial CLL

1.10.4.2. Genome-wide association studies in ALL

A GWAS in childhood ALL reported by Papaemmanuil *et al.* 2009, identified three risk alleles for childhood ALL, 7p12.2 (*IKZF1*, rs4132601), 10q21.2 (*ARID5B*, rs7089424) and 14q11.2 (*CEBPE*, rs2239633)(145). These authors also found that the 10q21.2 risk allele was specific for PBALL with hyperdiploidy and their data supports the hypothesis that common low-penetrance susceptibility alleles contribute to the risk of developing ALL(145). Trevino *et al.* 2009, also investigated two SNPs in *ARID5B* (rs10821936 and rs10994982) and found that these two SNPs were significantly associated with ALL. They concluded that germline variants contribute to the genetic susceptibility to ALL and to the cytogenetic subtype of ALL that develops(151). There is a well established ethnic difference in ALL between African-Americans and Caucasians which can be explained by the higher frequency

of the risk allele of rs10821936 in Caucasians compared to African-Americans (152). Healy *et al.* 2010, have replicated these SNPs in a French-Canadian cohort of 284 childhood ALL cases and 207 controls and confirmed five SNPs [rs7073837 (p=4.2 x 10^{-4}), rs10994982 (p=3.8 x 10^{-4}), rs10740055 (p=1.6 x 10^{-5}), rs10821936 (p=1.7 x 10^{-7}) and rs7089424 (p=3.6 x 10^{-7})] in the *ARID5B* gene with an increased susceptibility to childhood ALL.

1.10.4.3. Genome-wide association studies in therapy related AML

A side effect of cytotoxic therapy to treat a HM or other cancer is the risk of developing a HM as a result of the cytotoxic therapy effect on healthy cells. Therapy associated HM accounts for approximately 10-30% of cases of AML(153). Certain therapies are higher risk than others; however only a small proportion of patients will develop a therapy related AML, suggesting host factors also contribute to the risk of developing a therapy related HM. A recent GWAS found evidence of association of 3 SNPs for therapy related AML in patients with acquired loss of chromosome 5 or 7 or both(154).

1.10.4.4. Genome-wide association studies in mature B cell neoplasms

Skibola *et al.* 2009, undertook a GWAS of patients with NHL using the Illumina HumanHap550 beadchip in the hope of identifying specific SNP associations for the different subtypes of mature B cell neoplasms, in particular, follicular lymphoma, DLBCL and CLL(146). They found that one SNP (rs6457327) located on 6p21.33 was associated with susceptibility to follicular lymphoma(146). They concluded that further work is required to identify the causal variant(146)

1.11. *Whole genome sequencing in HMs*

With the publication of the human genome sequence by two groups in 2001 and the subsequent advancement of sequencing technology(155-159), whole genome sequencing of affected families' members and cancer genome is now possible at a reasonable price and timeframe. In a comparison of the genome sequence of leukaemic and skin cells from a patient with cytogenetically normal AML(160), ten acquired mutations were found in the leukemic cells. Two were in genes that have been implicated in leukaemia previously (*FLT3* and *NPM1*), but 8 were in novel genes(160), *CDH24*, *PCLKC*, *GPR123*, *EB12*, *PTPRT*, *KNDC1*, *SLC15A1* and *GRINL1B*(160).

1.12. *Discussion and Conclusion*

There is clear evidence that an inherited genetic abnormality can predispose to development of a HM, and that using families with multiple cases is a useful method for detecting genetic mutations. Due to the maturation process of haematological cells, a germline mutation is present in all the different types of HSC; then, if additional mutations occur, a different type of HM may develop depending upon how committed that HSC is. This has been shown with germline mutations in *RUNX1* and *CEBPA*, that different types of HMS can develop in these families. This is also supported by other acquired genetic changes in leukaemic cells that are not specific for any type of HM, but are seen in many different subtypes of HMs.

Mutations responsible for the familial risk for many other types of cancers (prostate cancer, breast cancer, colorectal cancer) have been identified by the use of familial studies; however, only one mutation in one familial CLL family has been found to

date. This has caused experts in the field to question the mode of inheritance in this condition and postulate that rather than of a Mendelian predisposition, a polygenic model is more likely(130). However it is clearly recognised in the literature that there is a genetic basis for the familial risk of HMs although genome-wide association studies of CLL(149) childhood ALL(145, 148) and NHL(146) have identified genetic variants accounting for only a small percentage of the heritable risk. Collections of large pedigrees with a range of HMs are rare but it is likely that the genetic study of such families will provide important insights into the genetic risk of familial HMs.

The hypothesis is that there is a genetic mutation that is segregating in some Tasmanian families resulting in an increased risk for developing a HM. The aims of this study are firstly to identify and recruit families with multiple affected members, secondly to confirm and classify the type of HM and locate pathology samples from the affected individuals and thirdly to identify the mutation in these families.

2. Chapter 2: Materials and Methods

2.1. *Patient recruitment from the 13 Tasmanian families and ethics approval*

This study has ethics approval by the Human Research Ethics Committee (Tasmania) network (H8551 from November 2005 for 4 years and has been extended for a further 4 years subject to annual progress reports being submitted). Ceri Flowers, a research assistant with the Menzies Research Institute, was responsible for contacting potential study participants. A survey and consent form (Appendix A) was sent to all adult living relatives of deceased individuals and affected individuals in the 13 families, when their doctor gave permission and a current address could be found. They were requested to provide information on a family history of a HM and consent to provide a peripheral blood or saliva sample. All participants provided written informed consent.

2.2. *DNA isolation*

DNA was extracted from ethylenediaminetetraacetic acid (EDTA) whole blood or from saliva using the Nucleon Bacc3 (Amersham Biosciences) and PureGene DNA isolation kits (Gentra Systems) respectively. Formalin fixed paraffin embedded tissue (FFPET) was collected from pathology laboratories (Royal Hobart Hospital, Hobart Pathology, North-West Pathology, Launceston Pathology, Bendigo Pathology and Newcastle Pathology) and the DNA extracted using the Pico Pure^{TM} DNA Extraction Kit (Arcturus Bioscience). Giemsa stained bone marrow slides were also collected from pathology laboratories (Royal Hobart Hospital, Hobart Pathology, North-West Pathology, Launceston Pathology) and the DNA extracted using the Pico Pure^{TM}

DNA Extraction Kit (Arcturus Bioscience). DNA from microdissected chromosomes was isolated using the Pico Pure™ DNA Extraction Kit (Arcturus Bioscience).

2.3. *DNA purification*

DNA from FFPET, Giemsa stained bone marrow aspirate smears or micro-dissected chromosomes were isolated as per 2.1. The DNA was then purified with the following method. The sample was made up to 200µl with water, then 100µl of phenol and 100µl of chloroform were added and the sample vortexed on high for 15 seconds. Supernatant was removed and put into a new tube. 1µl of glycogen, 10% of volume of sodium acetate (pH 5.2) and 2x volumes of 100% ethanol were added. The sample was mixed by inversion and put on ice for two hours. This was then centrifuged for 5 minutes at 250 RCF and the supernatant discarded. The pellet was washed in 400µl of 70% ethanol and centrifuged for 15 minutes at 250 RCF. The supernatant was discarded and the DNA air-dried and resuspended in water or TE buffer. Some samples were also put onto Whatman FTA cards (Interpath) and extracted as per manufacturers instructions.

2.4. *DNA whole genome amplification*

DNA from pathology samples and micro dissected chromosomes was amplified using genomiphi (GE Healthcare), Repli-g® whole genome amplification (Qiagen) and GenomePlex® (Sigma-Aldrich). The sample was then purified using Agencourt AMPure (Beckman Coulter genomics).

2.5. *Assessing DNA amount and quality*

The amount and molecular weight of the DNA extracted was assessed on a 2% agarose gel. The concentration was also assessed using a Nanodrop (Thermo Scientific) and the concentration of DNA and the 260/280 and 230/260 ratios were recorded.

2.6. *Microsatellites*

Microsatellites were selected using the UCSC genome browser (http://genome.ucsc.edu). The region of interest was inserted into the genome browser and the microsatellites within the region of interest were assessed for their suitability (ie size, heterozygosity score). The recommended forward and reverse primer sequences were recorded and also submitted to "BLAT" (http://genome.ucsc.edu/cgi-bin/hgBlat?command=start) to ensure their uniqueness. Primers were order from Geneworks**.** The forward primer was labelled with a fluorchrom (Hex or Fam). Each primer was made up to 100µM concentration with water. A working solution of both the forward and reverse primer was made up to a concentration of 10µM by adding 5µl of the forward primer and 5ul of the reverse primer to 90µl of water. Each set of primers was optimised for an annealing temperature. The size of the microsatellite in each person was assessed using an ABI PRISM 310 Genetic Analyzer (Applied Biosystems). Raw traces were formatted in GeneScan (Applied Biosystems) and then imported into Genotyper (Applied Biosystems) for analysis. Microsatellite allele frequencies were estimated from 25 unrelated Tasmanian people who did not have a haematological malignancy. Primer sequences are listed in Appendix B.

2.7. *Single Nucleotide Polymorphisms (SNPs)*

SNPs were chosen from the UCSC genome browser (http://genome.ucsc.edu) or from the affymetrix SNP array. SNP's were chosen where LK16-1 was homozygous for the rare allele. The DNA sequence for 200bp either side of the SNP was obtained from the UCSC genome browser and primers were designed for this region using Primer3 (http://frodo.wi.mit.edu/primer3/input.htm). The primer sequence were put through "BLAT" (http://genome.ucsc.edu/cgi-bin/hgBlat?command=start) to ensure their uniqueness and ordered from Geneworks**.** Each primer was made up to 200µM concentration with water. A working solution of both the left and right primer was made up to a concentration of 5µM by adding 5µl of the left primer and 5ul of the right primer to 90µl of water. Each set of primers was optimised for an annealing temperature. PCR products were cleaned with Ampure (Beckman Coulter Genomics) as per manufactures instructions and the sequencing reaction run using Big Dye Terminator (BDT) 3.1 sequencing enzyme (Applied Biosystems) in the following concentration: DNA 1µl per 100bp to be sequenced, 0.25µl BDT sequencing enzyme (Applied Biosystems), 1.75µl of Sequencing Buffer (Applied Biosystems), 1µl of forward or reverse primer and water to make up to 10µl. Samples underwent 25 cycles at 96^{o}C for 10 seconds, 50 ^{o}C for 5 seconds and 60^{o}C for 4 minutes. The sequencing reaction was then purified using CLEANSEQ (Beckman-Coulter) as per manufactures instructions and 30µl per sample run on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems) and analysis of the raw traces complied and analysed with Sequencer 4.10.1 (Gene codes Corporation). Primer sequences are listed in Appendix B.

2.8. *Sequencing of genes*

The DNA sequences of genes to be sequence were obtained from the UCSC genome browser (http://genome.ucsc.edu). Primers were then designed using the ExonPrimer for the entire cDNA, 5'UTR, 3'UTR, splice sites and 500bps upstream and downstream of the gene. The primers generated by this program were checked in "BLAT" for uniqueness and ordered from Geneworks. These primers were treated in the same manner as the primers used in sequencing the SNPs (Chapter 2.7). Primer sequences are listed in Appendix B.

2.9. *Conventional Cytogenetics*

Whole peripheral blood was collected into a 9ml Lithium heparin peripheral blood tube. 0.5mls of peripheral blood was then added to 7ml of RPMI 1640 media (Sigma-Aldrich) supplemented with 0.8% glutamax (Gibco, Invitrogen) and 13% fetal calf serum (FCS) (Gibco, Invitrogen). 0.4mls of phytohaemagglutin (PHA) (Roche) and 0.1ml of antibiotics (Gibco, Invitrogen) were added and the culture was incubated for 48 hours at 37^{o}C. 100µl of thymidine (Sigma-Aldrich) was added and 18 hours later this block was released with 0.1ml of 2-deoxycytidine hydrochloride (Sigma-Aldrich). The culture was then incubated at 37^{o}C for 3 hours and forty-five minutes then 0.1µg/ml of colcemid (Gibco, Invitrogen) was added and incubated at 37^{o}C for 20 minutes. The culture was then centrifuged at 250 RCF for 6 minutes. The supernatant is discarded and the pellet was resuspend in 8ml of 0.038mol/L potassium chloride and incubated for 27 minutes at room temperature. Then 1ml of fixative was added, mixed and centrifuge at 250 RCF for 6 minutes. The supernatant is discarded and the pellet resuspend in 6mls of fixative and centrifuge at 250 RCF

for 6 minutes this step repeated. The slides were made, aged for 2 days and the chromosomes giemsa trypsin leishman (GTL) banded and analysed using IKAROS software (Metasystems, Germany) and the results were reported according to international system of cytogenetic nomenclature (ISCN)(161).

2.10. *Molecular cytogenetics*

FISH probes were ordered from The Hospital for Sick Children, Canada to cover the microsatellite regions on chromosome 15 and a region on chromosome 4q13. FISH probes were ordered from Vysis (Abbott Molecular, Australia) for *IGH@*, 13q14 (D13S319), 13q34, CEP(12), *TP53* and *ATM* and used according to product insert. Pretreatment of FFPET samples was with the Vysis paraffin pre-treatment reagent kit III (Vysis Abbott Molecular). FISH was analysed using an Axioskop 2plus microscope using ISIS software (Metasystems, Germany) and results reported according to ISCN(161).

2.11. *Cell Lines*

Epstein barr virus (EBV) containing supernatant was obtained from culturing monkey B95-8 cells. The lymphocytes from subjects were purified with a ficoll-Paque^{TM} (GE Healthcare) gradient, washed in RPMI (Sigma-Aldrich) and resuspended in 8mls of medium (RPMI with 13% FCS and 0.8% glutamax) with 2ug/ml cyclosporine (Sigma-Aldrich). This was split into two T25 falcon flasks, 4ml/flask and 1ml EBV supernatant was then added to each flask. The cultures were fed weekly for the first two weeks and then bi-weekly and checked for cell growth

(clump formation). The cultures were split when there was sufficient cell growth (mostly after 4 weeks). RNA was then extracted using PureLink™ (Invitrogen) and some cultures were stored for DNA extraction on Whatman FTA cards (Interpath). The remaining cultures were frozen at 10 x 10^{-6} cells in 2ml RPMI supplemented with 15% FCS and 5% DMSO.

2.12. *Laser Dissection of chromosomes*

Chromosome preparations were made as described in Chapter 2.9. Several slides were made on each sample. Slides were aged for 2 days at 60^{O}C and then GTL banded. The chromosome region of interest was laser micro-dissected using a PALM MicroBeam (Carl Zeiss) as per the manufactures instructions into a cap tube containing buffer from the Pico Pure^{TM} DNA Extraction Kit (Arcturus Bioscience).

2.13. *Standard PCR conditions*

The PCR mix consisted of 5ul of Gotaq® Green (Promega), 0.4µl of primer set, 3.6µl of water and 1µl of DNA at 10ng/µl. The standard PCR cycling conditions were denaturation at 94^{O}C for 5 minutes, followed by 30 cycles at 94^{O}C for 30 seconds, specific annealing temperature 60^{O}C for 30 seconds and 72^{O}C for 30 seconds. The reaction was cooled to 10^{O}C for 10 minutes. The annealing temperature for each PCR was optimised on a gradient cycle. 30 cycles was standard for all PCRs, except when the source of the DNA was from FFPET or giemsa bone marrow aspirate smears, then 35 cycles were performed.

3. Chapter 3: The clinical aspects of the 13 families with HMs.

3.1. *Introduction*

A positive family history is a well-documented risk factor for HMs. Reported collections of large families with multiple cases of varied types of HM are relatively rare(19), with most families in the literature having the same disease, mostly CLL(6-8, 15, 18, 19, 24), however other subtypes (myeloma, AML, MPN and lymphomas) have been described(5, 9, 10, 13, 14, 17, 23). Many of the families in the literature are not complete and details are missing in terms of confirmation of disease, clinical features or age at diagnosis.

A collection of 13 families featuring a dense aggregation of HM cases, was ascertained from a population based case control study conducted between 1972-1980 in Tasmania(33, 112, 113) as described in Chapter 1.10.1. The original records of this study have been retained (by Jean Panton) and were available for review by myself. They noted that the "risk for children of parents is higher than the risk to siblings and that the risk of siblings is higher than that to parents" (33).

3.2. *Materials and Methods*

3.2.1. Updating of the families from the original 1970's Tasmanian study

The families from the original study(33, 112, 113) were reviewed and current generations included and cross-referenced with the Tasmanian Cancer Registry (TCR) and the genealogical databases of the Menzies Research Institute. The TCR was commenced in 1982 with mandatory reporting for doctors, including HMs.

Formerly, not all HMs were recorded by the TCR. Waldenstroms macroglobulinaemia, some MPNs and MDS were not recorded as they were considered to be "pre-malignant" conditions not cancers; however with the current WHO classification system(1) these disorders are now registered. The families prioritised for further study were selected based upon the number of cases affected, multiple generations affected, or sibling pairs affected.

3.2.2. Survey of living patients from the TCR

Ceri Flowers, a research assistant with the Menzies Research Institute, conducted this. A list of patients who were registered with the TCR as having a HM and living as of the 1st of June 2006 was generated by the TCR. An information and consent form (Appendix A) was sent to all living affected cases with a HM when their doctor gave permission and a current address could be found. They were requested to provide information on a family history of a HM.

3.2.3. Confirmation of diagnosis

Medical records from the Royal Hobart Hospital (RHH), pathology records from the Pathology Department of the RHH, flow cytometry records from the University of Tasmania's Oncology and Immunology Laboratory, files from the 1970's study (33, 112, 113) and records of the TCR were obtained and pathology samples located and reviewed by myself to reconfirm the type of HM. If one or more of a pathology sample, report, doctors' correspondence or death certificate was found and confirmed the HM, then that person was counted as a reconfirmed case. If records no longer existed or they had a HM that was not recorded by the TCR (such as Waldenstroms

macroglobulinaemia, MPN or MDS), but was recorded as affected in the original study (33, 112, 113), then that person would be considered as a confirmed case.

3.2.4. Statistical analysis

Graphs were drawn with Microsoft® Excel® (Mircosoft Corporation). Pedigrees were generated using Smartdraw (Smartdraw Software LLC). Student T-tests were generated with Microsoft® Excel® (Mircosoft Corporation). Age at diagnosis was calculated using Microsoft® Excel® (Mircosoft Corporation), ((Date of diagnosis – Date of birth)/365) and rounded to the next full year. Test for trend was generated using GraphPad Prism software version 5.

3.3. *Results: Clinical features of the 13 families*

3.3.1. Leukaemia Family 1 (LK1)

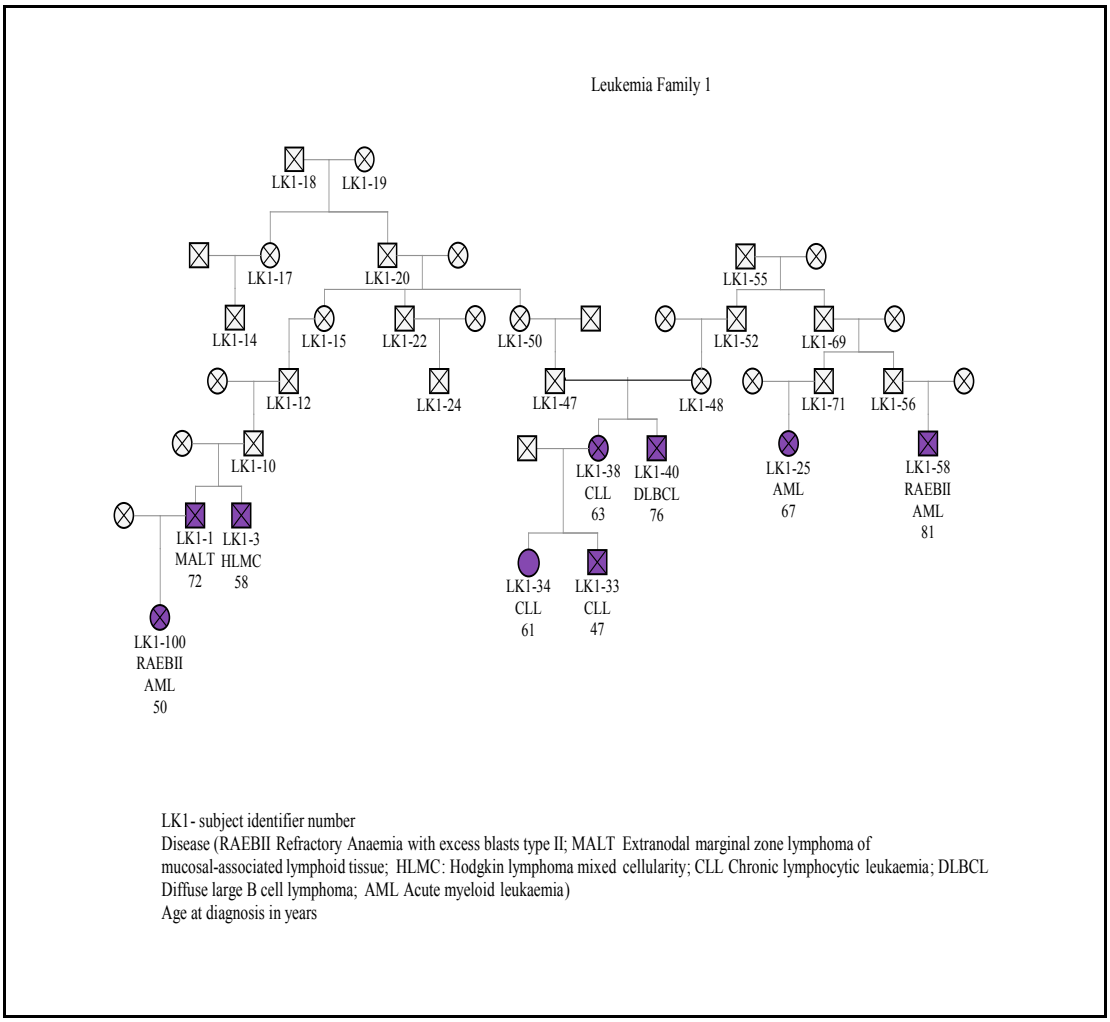
There are 9 cases in this family. The disease characteristics are given in Table 3.3.1.1. All cases in this family could be reconfirmed. An abbreviated pedigree showing the relationship of the affected persons is given in Figure 3.3.1.1. It should be noted that only 3 of these cases were in the original study(33, 112, 113), and that a further 6 subjects in this family have been diagnosed since then. All pathology records could be traced for this family. The male to female ratio was 5:4. The median age at diagnosis was 63 years (range 47-81years). There were two cases of early onset disease, notably CLL in a 47 year old and MDS in a 50 year old. There was one case of therapy-related MDS developing in a patient who originally had DLBCL (LK1-40). The MDS developed 14 months after the diagnosis of DLBCL. The two cases of primary MDS transformed to AML within one year of diagnosis. One patient who had HL was found at post mortem to also have an adenocarcinoma of his right upper lobe of his lung. There was also a cluster of cases of CLL, with a mother (LK1-38) and two of her children affected with CLL (LK1-34 and LK1-33).

Table 3.3.1.1: The disease characteristics of LK1.

ID	Gender	DOB	Date of Diagnosis	Date of Death	Diagnosis*	Age at diagnosis
001	M	07/12/1926	01/05/1999	15/02/2007	MALT	72
003	M	13/12/1936	01/12/1994	05/02/1995	HL MC	58
025	F	13/04/1906	01/11/1972	03/01/1973	AML Monocytic	67
033	M	09/08/1928	01/09/1975	04/02/1979	CLL	47
038	F	21/11/1907	01/02/1971	28/12/1976	CLL	63
040	M	01/01/1912	01/08/1987	09/11/1988	DLBCL	76
058	M	01/10/1910	01/04/1991	27/04/1991	RAEB AML	81
034	F	14/11/1930	01/01/1991	Alive	CLL	61
100	F	05/06/1956	01/01/2006	27/12/2006	RAEBII AML	50

**are defined in the abbreviation section*

Figure 3.3.1.1 Pedigree of LK1



3.3.2. Leukaemia Family 2 (LK2)

There are 12 cases in this family (Figure 3.3.2.1). The disease characteristics are given in Table 3.3.2.1. Eight cases in this family could be reconfirmed and one case was excluded. Six cases were from the original study(33, 112, 113). The median age at diagnosis in this family was 65 years (range 39-83 years). The male to female ratio was 10 to 4. One patient also developed breast cancer 31 years after the diagnosis and successful treatment of a lymphoma. LK2-33 transformed from CMML to AML 13 months after diagnosis.

Figure 3.3.2.1 Pedigree of LK2

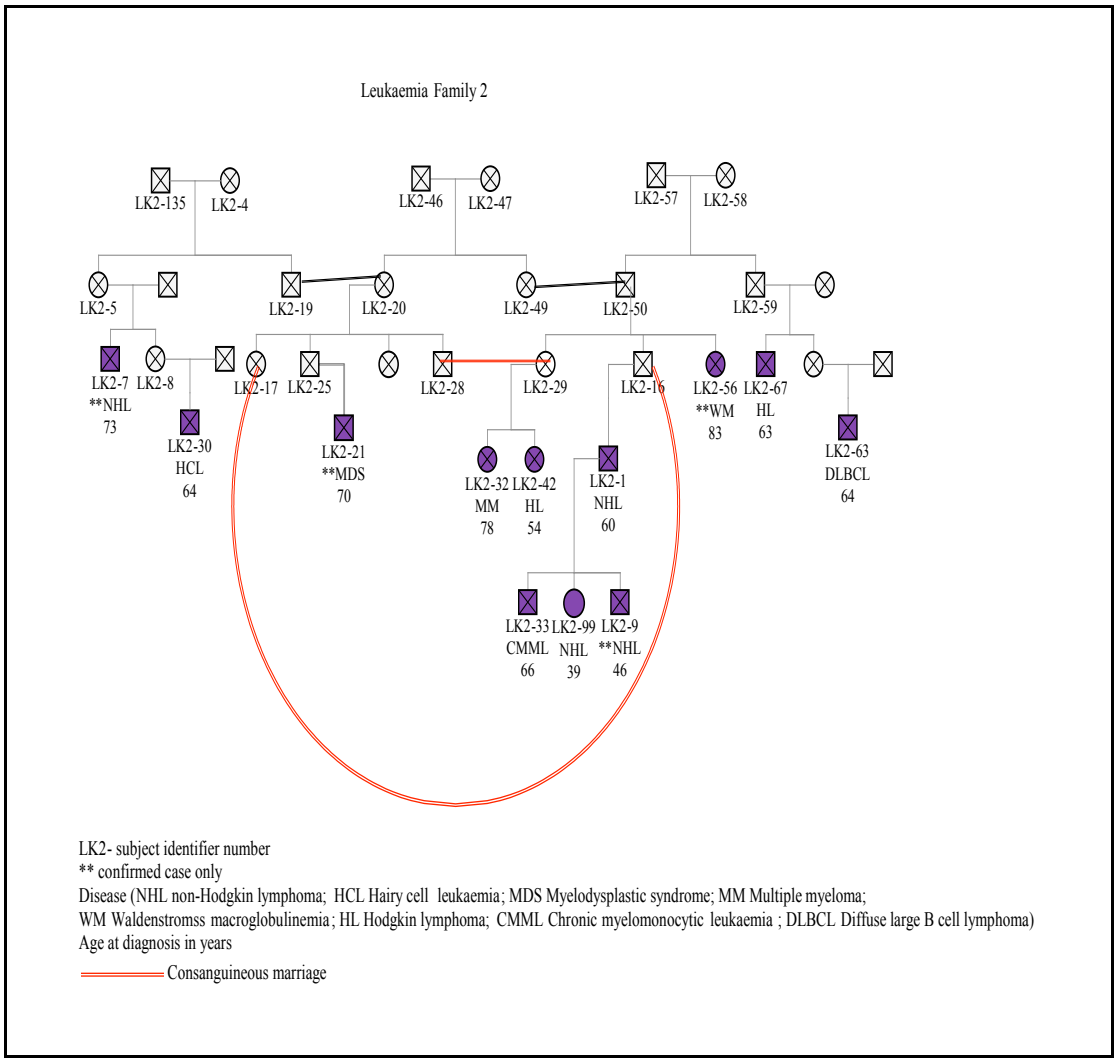


Table 3.3.2.1: The disease characteristics of LK2.

ID	Gender	DOB	Date of Diagnosis	Date Of Death	Diagnosis*	Age at Diagnosis
001	M	12/12/1907	01/01/1968	08/08/1968	NHL	60
007	M	29/12/1882	01/05/1954	19/01/1955	**NHL	73
009	M	1 /12/1939	01/01/1985	10/01/1993	**NHL	46
021	M	8 /06/1913	01/12/1985	20/04/1993	**MDS	72
030	M	15/08/1915	01/05/1979	10/03/1979	HCL	64
032	F	1 /12/1903	01/07/1980	23/12/1986	MM	77
033	M	9 /02/1928	01/07/1994	10/03/1996	CMML AML	66
042	F	1 /10/1915	01/10/1968	14/03/1974	HL	53
056	F	6 /01/1900	01/01/1983	25/06/1984	**WM	83
063	M	29/07/1931	01/07/1995	24/07/2001	DLBCL	64
067	M	18/09/1910	01/01/1972	10/06/1973	HL	62
099	F	17/07/1936	01/06/1975	Alive	NHL	39

*are defined in the abbreviation section

** confirmed cases

3.3.3. Leukaemia Family 4 (LK4)

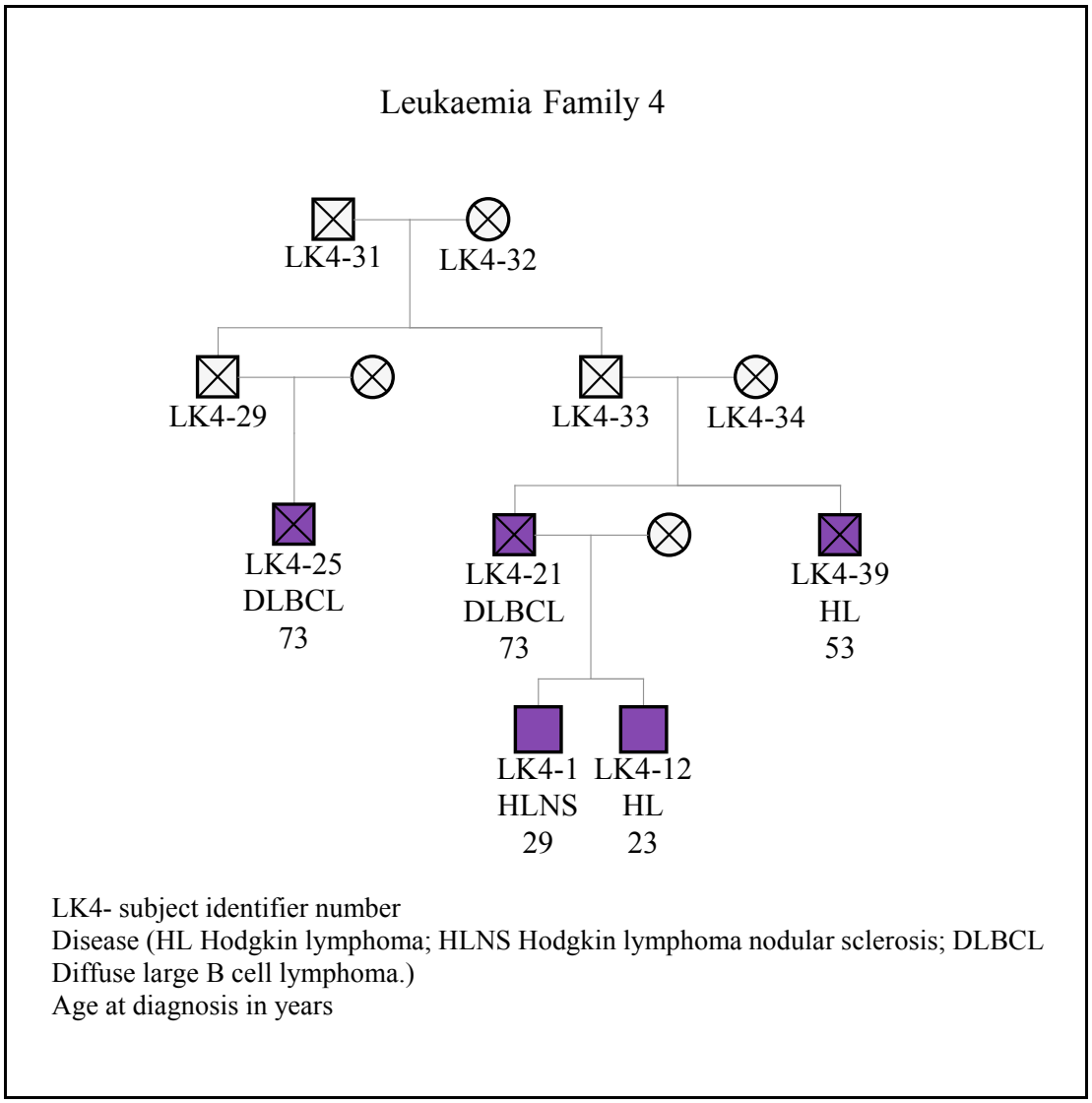
There are 5 cases in this family (shown in Figure 3.3.3.1). The disease characteristics are given in Table 3.3.3.1. All cases could be reconfirmed in this family. Four cases were from the original study(33, 112, 113). All cases in this family are male and the majority of cases were Hodgkin lymphoma (3 cases). The other two cases were DLBCL. The median age at diagnosis was 53 years (range 23-73 years). LK4-25 was also diagnosed with non-invasive papillary transitional cell carcinoma grade 1 of the bladder, 12 years after the diagnosis of DLBCL.

Table 3.3.3.1: The disease characteristics of LK4 family

ID	Gender	DOB	Date of Diagnosis	Date Of Death	Diagnosis*	Age at diagnosis
1	M	6 /09/1941	12/04/1970	Alive	HL NS	29
12	M	14/01/1947	01/04/1970	Alive	HL stage IIIB	23
21	M	3 /10/1906	01/07/1979	28/11/1980	DLBCL	73
25	M	31/10/1914	01/10/1987	17/10/2001	DLBCL	73
39	M	12/01/1912	1965	25/06/1965	HL	53

*are defined in the abbreviation section

Figure 3.3.3.1 Pedigree of LK4



LK4- subject identifier number
Disease (HL Hodgkin lymphoma; HLNS Hodgkin lymphoma nodular sclerosis; DLBCL Diffuse large B cell lymphoma.)
Age at diagnosis in years

3.3.4. Leukaemia Family 16 (LK16)

There are 12 cases in this family. An abbreviated pedigree is shown in Figure 3.3.4.1. The disease characteristics are given in Table 3.3.4.1. All 12 cases in this family could be reconfirmed. An additional case was excluded, as review of his medical records was not in keeping with a HM; his pancytopenia was due to chemotherapy for his bowel cancer. Only 4 cases were from the original study(33, 112, 113). The male to female ratio is 10 to 2. The median age at diagnosis was 65 years (range 35-80 years). This is an unusual family in that it has a small cluster of five siblings out of seven affected with a HM. Four of the siblings had CLL and the fifth had DLBCL. Also, the mother of these siblings was reported to have died from CML. Her death certificate has been traced and the cause of death was listed as CML in 1943. There are also 7 other affected subjects in this family with the most recent being diagnosed in January 2008 with MM. Two of the affected also had another type of cancer; LK16-14 was diagnosed with melanoma and LK16-108 with prostate cancer. LK16-101 developed AML 12 years after treatment for DLBCL.

Figure 3.3.4.1 Pedigree of LK16

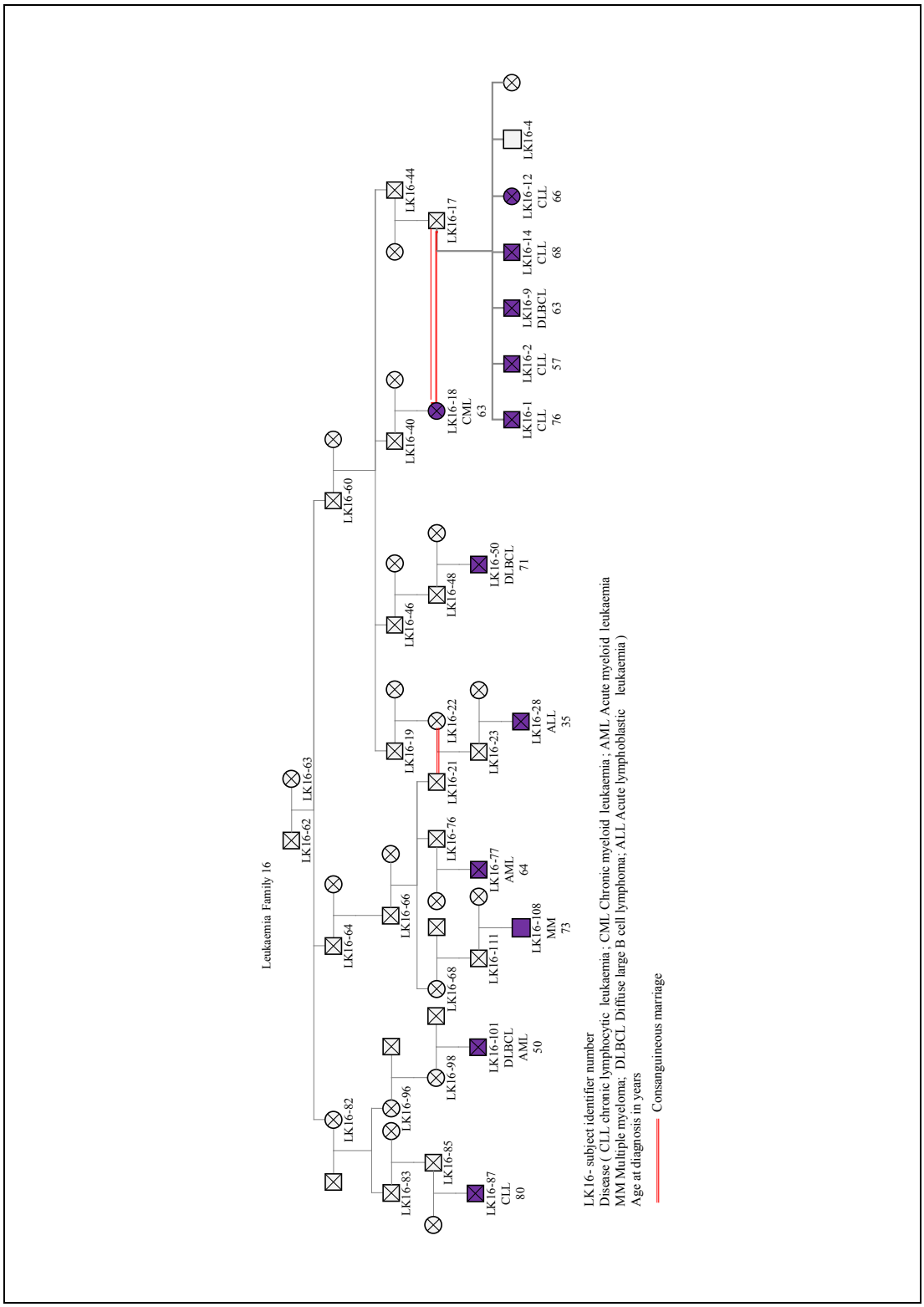


Table 3.3.4.1: Disease characteristics of LK16 family

ID	Gender	DOB	Date of Diagnosis	Date Of Death	Diagnosis*	Age at diagnosis
001	M	05/03/1921	01/05/1997	17/08/2006	CLL	76
002	M	11/11/1922	13/02/1980	19/01/1986	CLL	57
009	M	23/02/1919	01/07/1982	6 /01/1995	DLBCL	63
012	F	07/06/1914	01/07/1980	23/01/1999	CLL	66
014	M	02/05/1910	25/01/1978	1 /02/1982	CLL	68
018	F	17/10/1880	13/9/1943	13/09/1943	CML	63
028	M	01/10/1937	01/05/1972	14/09/1972	ALL	35
050	M	04/09/1915	01/10/1986	18/03/1987	DLBCL	71
077	M	04/12/1911	29/09/1975	29/09/1975	AML M1	64
087	M	05/11/1906	01/05/1986	13/07/1986	CLL	80
101	M	02/11/1930	01/09/1980	17/06/1998	DLBCL AML	50
108	M	15/09/1935	3/01/2008	Alive	MM	73

*are defined in the abbreviation section

3.3.5. Leukaemia Family 26 (LK26)

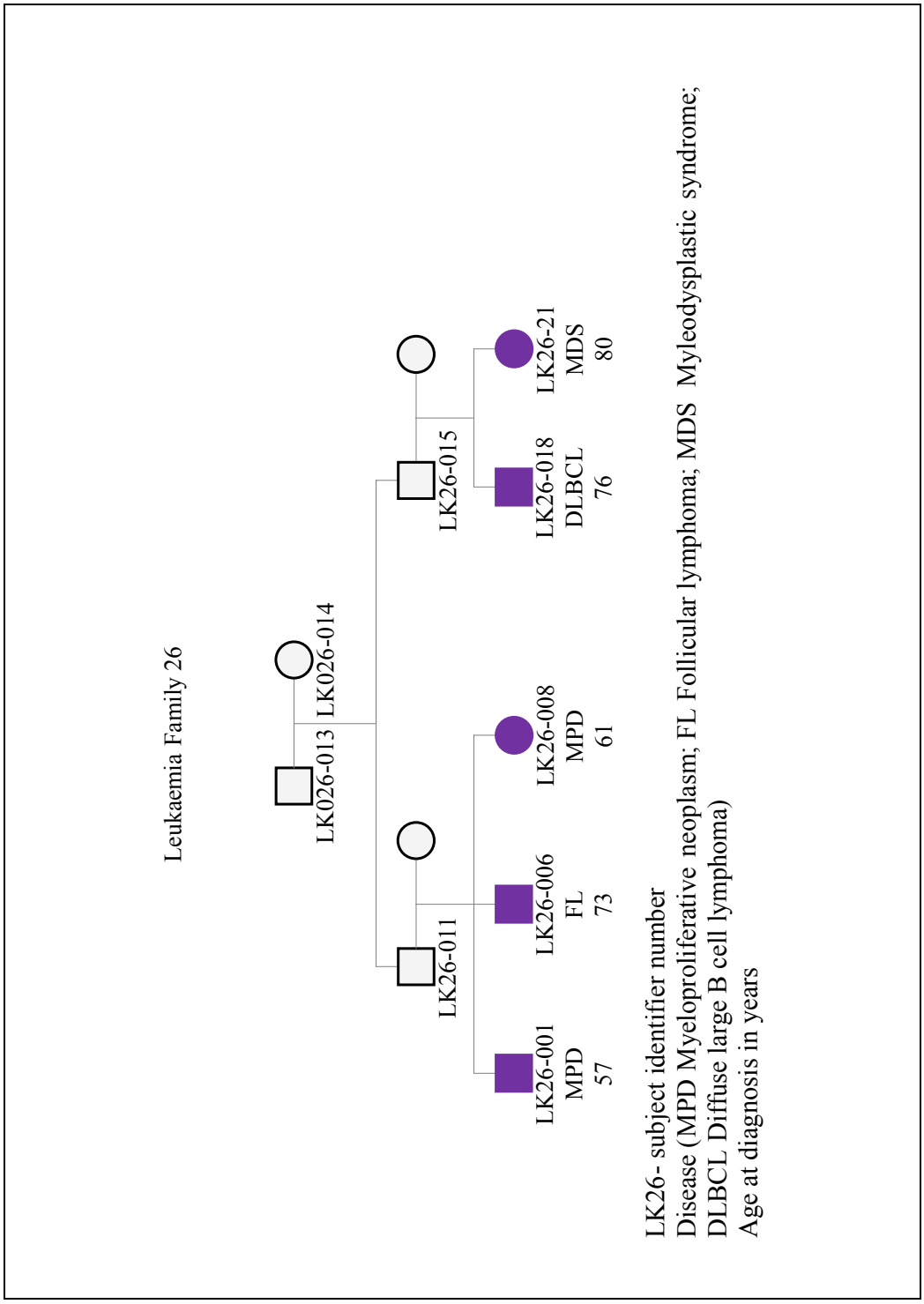
There are 5 cases in this family. An abbreviated pedigree in shown in Figure 3.3.5.1. The disease characteristics are given in Table 3.3.5.1. All cases could be reconfirmed in this family. The male to female ratio was 3 to 2. The median age at diagnosis was 69.4 years (range 57-80). Since the original study(33, 112, 113), two more cases have been diagnosed in this family. The one case of a MDS (LK26-21) transformed to AML within one year of diagnosis.

Table 3.3.5.1: Disease characteristics of LK26 family

ID	Gender	DOB	Date of Diagnosis	Date Of Death	Diagnosis*	Age at diagnosis
001	M	1 /06/1904	21/11/1960	20/09/1972	PV	56
006	M	16/04/1908	01/06/1981	7 /03/1996	FL	73
008	F	27/06/1911	01/07/1972	6 /04/1979	PV	61
018	F	13/12/1897	01/06/1973	22/10/1974	DLBCL	76
021	M	20/11/1910	01/11/1990	8 /12/1990	MDS AML	80

*are defined in the abbreviation section

Figure 3.3.5.1 Pedigree of LK26



3.3.6. Leukaemia Family 51 (LK51)

There are 15 cases in this family. An abbreviated pedigree is shown in Figure 3.3.6.1. The disease characteristics are given in Table 3.3.6.1. Ten cases could be reconfirmed in this family. Seven cases were from the original study(33, 112, 113). The male to female ratio was 10 to 5. The median age at diagnosis was 56 years (range 6-81 years). There were 2 cases of childhood onset in this family. Interestingly neither had ALL; one had Burkitt lymphoma and the second had AML.

Figure 3.3.6.1 Pedigree of LK51

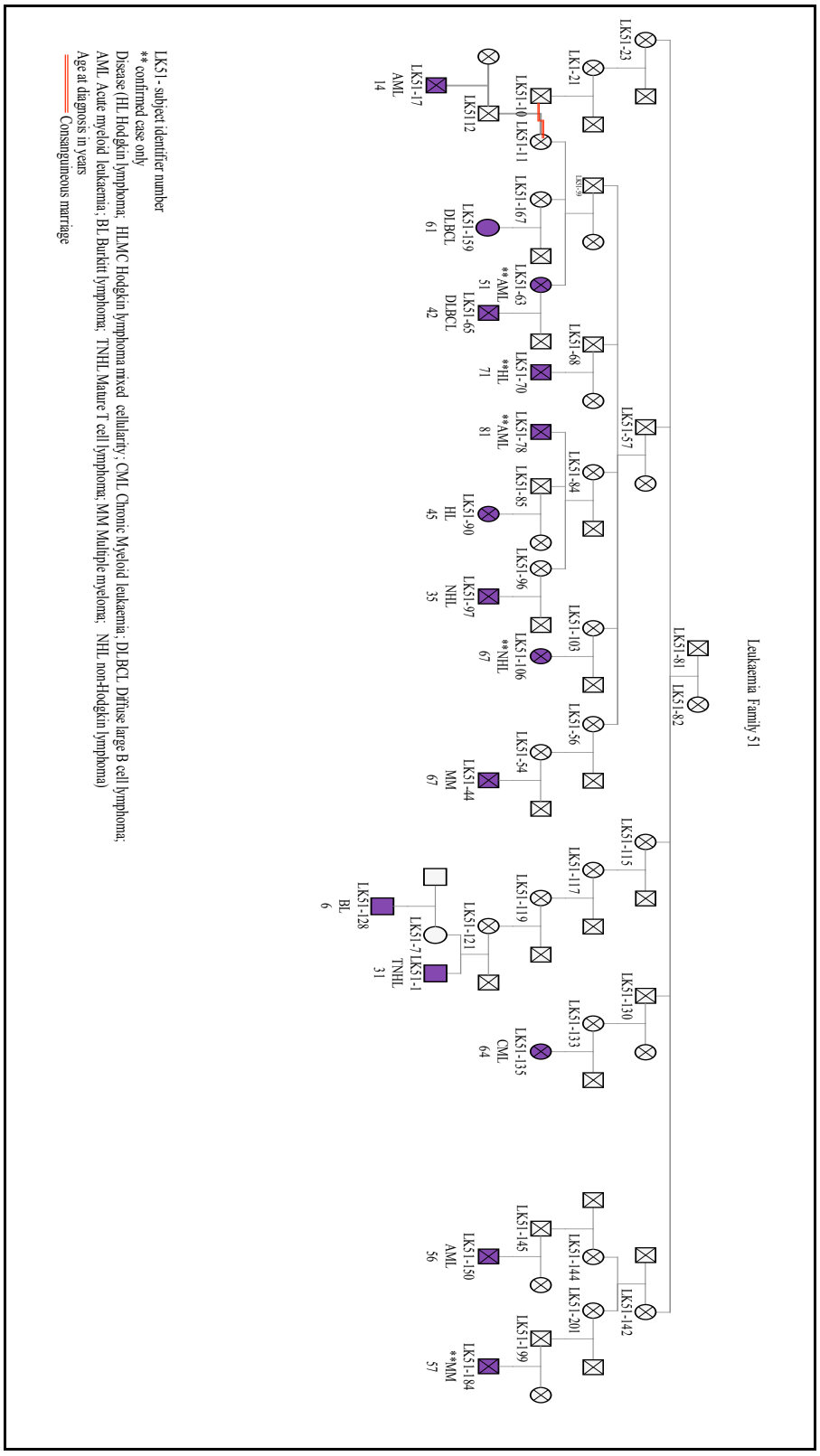


Table 3.3.6.1: The disease characteristics of LK51

ID	Gender	DOB	Date of diagnosis	Date of Death	Diagnosis*	Age at diagnosis
001	M	01/06/1967	01/04/1998	Alive	TNHL	31
017	M	24/08/1959	01/05/1973	01/07/1973	AML	14
044	M	25/11/1910	01/05/1977	04/06/1977	MM	67
063	F	01/08/1921	01/03/1972	30/05/1972	**AML	51
065	M	18/07/1942	01/08/1984	30/09/1985	DLBCL	42
070	M	19/01/1905	01/03/1976	02/10/1977	**HL	71
078	M	31/07/1899	01/05/1980	01/06/1980	**AL	81
090	F	17/01/1927	01/04/1972	24/01/1984	HL	45
097	M	15/04/1948	01/11/1983	14/11/1988	NHL	35
106	F	28/01/1904	01/10/1971	18/07/1995	**NHL	67
128	M	30/01/1996	19/11/2002	Alive	BL	6
135	F	01/03/1906	01/06/1970	24/10/1980	CML	64
150	M	26/02/1932	01/12/1988	18/05/1989	AML (M1)	56
159	F	15/04/1938	07/01/1997	Alive	DLBCL	59
184	M	15/07/1935	01/09/1992	25/08/1993	**MM	57

*are defined in the abbreviation section

** confirmed cases

3.3.7. Leukaemia Family 65 (LK65)

There are 8 cases in this family. An abbreviated pedigree is shown in Figure 3.3.7.1. The disease characteristics are given in Table 3.3.7.1. Five cases could be reconfirmed in this family. Five cases were from the original study(33, 112, 113). There is a reversed male to female ratio in this family (3 to 5). The median age at diagnosis was 59 years (range 14-79 years). Also of note in this family, is that the child (LK65-39) that was affected had AML not ALL, as ALL is more common in childhood then AML.

Figure 3.3.7.1 Pedigree of LK65

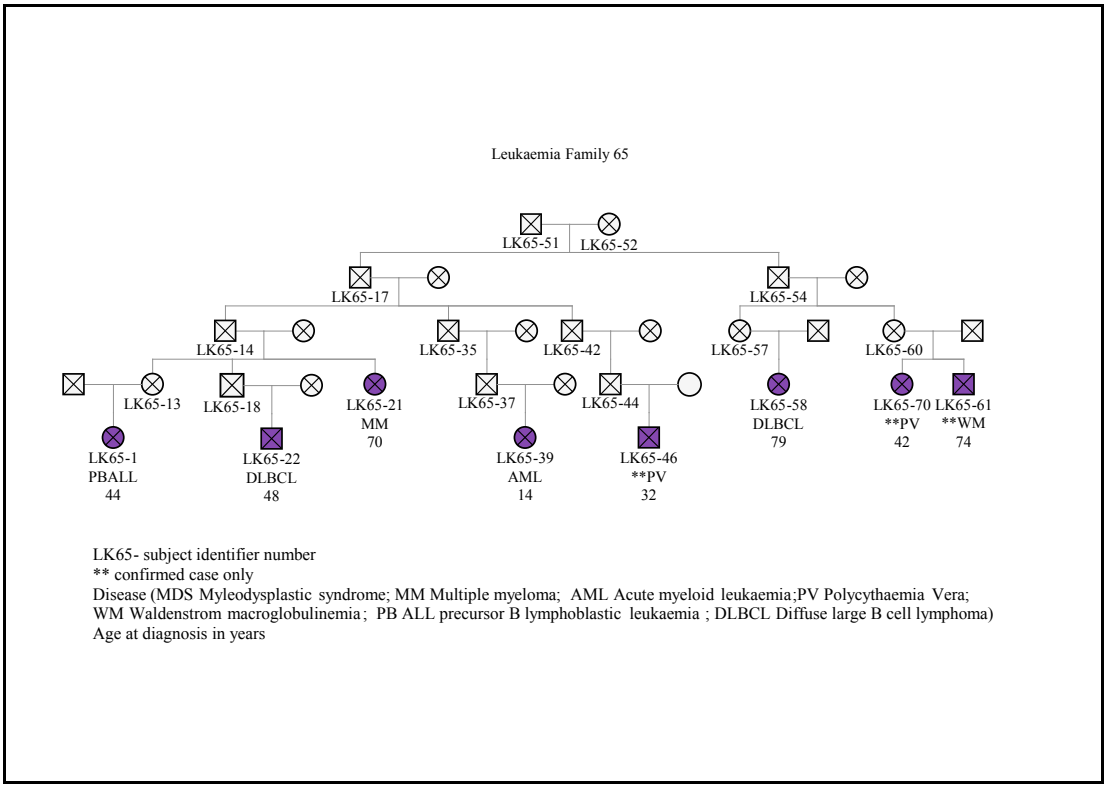


Table 3.3.7.1: The patient characteristic of LK65.

ID	Gender	DOB	Date of diagnosis	Date Of Death	Diagnosis*	Age at diagnosis
001	F	01/01/1932	01/02/1976	25/10/1976	PBALL	44
021	F	06/03/1907	01/03/1977	15/04/1977	MM	70
022	M	14/12/1934	01/03/1983	6 /11/1983	DLBCL	48
039	F	27/08/1951	01/03/1965	10/02/1966	AML M0	14
046	M	27/08/1946	01/04/1978	28/12/1995	PV**	32
058	F	22/05/1900	01/11/1979	11/03/1981	DLBCL	79
061	M	12/09/1909	01/06/1983	26/09/1992	WM**	74
070	F	11/03/1921	01/02/1963	03/07/1972	PV**	42

*are defined in the abbreviation section

** confirmed cases

3.3.8. Leukaemia Family 124 (LK124)

There are 18 cases in this family. An abbreviated pedigree is shown in Figure 3.3.8.1. The disease characteristics are given in Table 3.3.8.1. All cases in this family could be reconfirmed. Only 3 cases were from the original study(33, 112, 113). The male to female ratio was 15 to 3. The median age at diagnosis was 70 years (range 30-84 years). While confirming the diagnosis and classification, of the diagnosis in this family, some unusual clinical features emerged, namely, that several of the patients had disease at unusual anatomical sites. There were three cases of primary central nervous system (CNS) lymphoma (all DLBCL), two involved brain, and one spinal cord. One of these cases also developed a bone lesion in their rib 6 months after diagnosis. A fourth case of DLBCL also had an unusual site of presentation with primary bone lymphoma. There was also a case of rapid progression of a solitary soft tissue plasmacytoma progressing to multiple myeloma within 11 months of diagnosis. This family also had two cases of primary DLBCL presenting as gastric lymphoma. Three of the affected also had other cancers (LK124-1 and LK124-194 both with lung carcinoma and LK124-202 with testicular carcinoma).

Figure 3.3.8.1 Pedigree of LK124

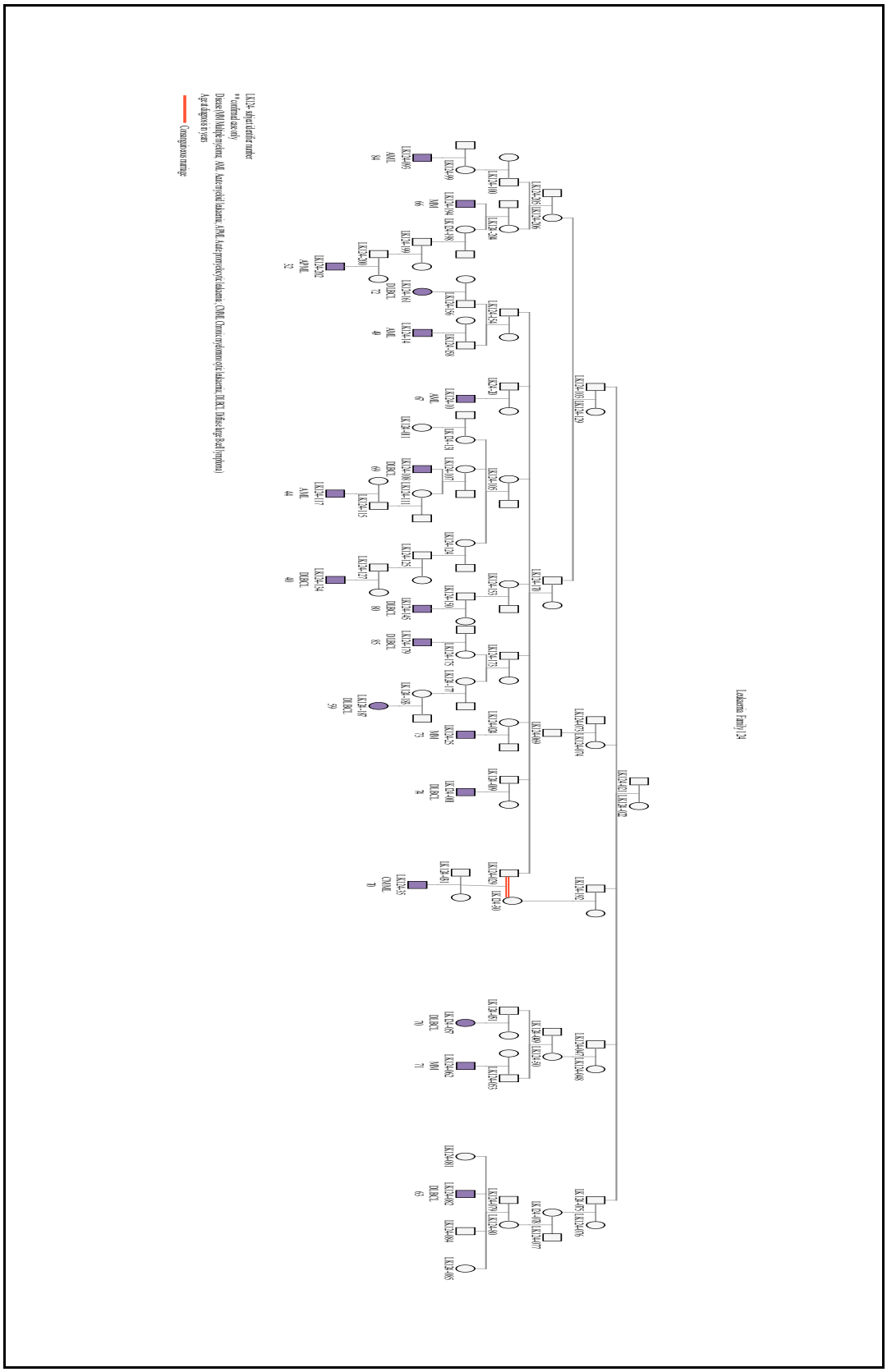


Table 3.3.8.1: The disease characteristics of LK124

ID	Gender	DOB	Date of Diagnosis	Date of Death	Diagnosis *	Age at diagnosis
1	M	08/08/1923	01/02/1997	17/04/2000	DLBCL	74
10	M	25/05/1914	01/02/1981	07/03/1981	AML M6	67
14	M	6 /03/1931	01/12/1979	01/02/1982	AML M1	49
25	M	21/02/1901	01/01/1974	06/12/1974	MM	73
35	M	20/02/1929	27/08/1998	26/10/2000	CMML	70
57	F	7 /11/1922	01/05/1992	19/05/1993	DLBCL	70
62	M	13/11/1923	01/09/1994	29/10/1996	MM	71
82	M	11/04/1931	01/07/1994	20/03/1995	DLBCL	63
93	M	29/12/1916	01/12/2000	12/12/2000	AML NOS	84
108	M	01/12/1917	01/08/1986	03/09/1986	DLBCL	69
117	M	08/07/1951	01/09/1995	Alive	AML	44
134	M	22/06/1959	01/09/1999	Alive	DLBCL	40
145	M	27/02/1920	01/05/2000	30/12/2000	DLBCL	80
161	F	14/02/1926	01/08/1998	22/09/1998	DLBCL	73
179	M	19/08/1919	02/06/2004	Alive	DLBCL	85
187	F	14/12/1943	01/01/2003	17/10/2006	DLBCL	59
194	M	10/07/1903	01/10/1969	04/03/1977	Plasma -cytoma MM	66
202	M	19/01/1972	25/02/2002	Alive	APML	30

*are defined in the abbreviation section

3.3.9. Leukaemia Family 132 (LK132)

There are 5 cases in this family. An abbreviated pedigree is shown in Figure 3.3.9.1. The disease characteristics are given in Table 3.3.9.1. Only 2 cases could be reconfirmed in this family. Only 1 case was from the original study(33, 112, 113). This is a smaller family with a cluster of affected subjects. The male to female ratio was 4 to 1. The median age at diagnosis was 59 years (range 49 to 72 years).

Figure 3.3.9.1 Pedigree of LK132

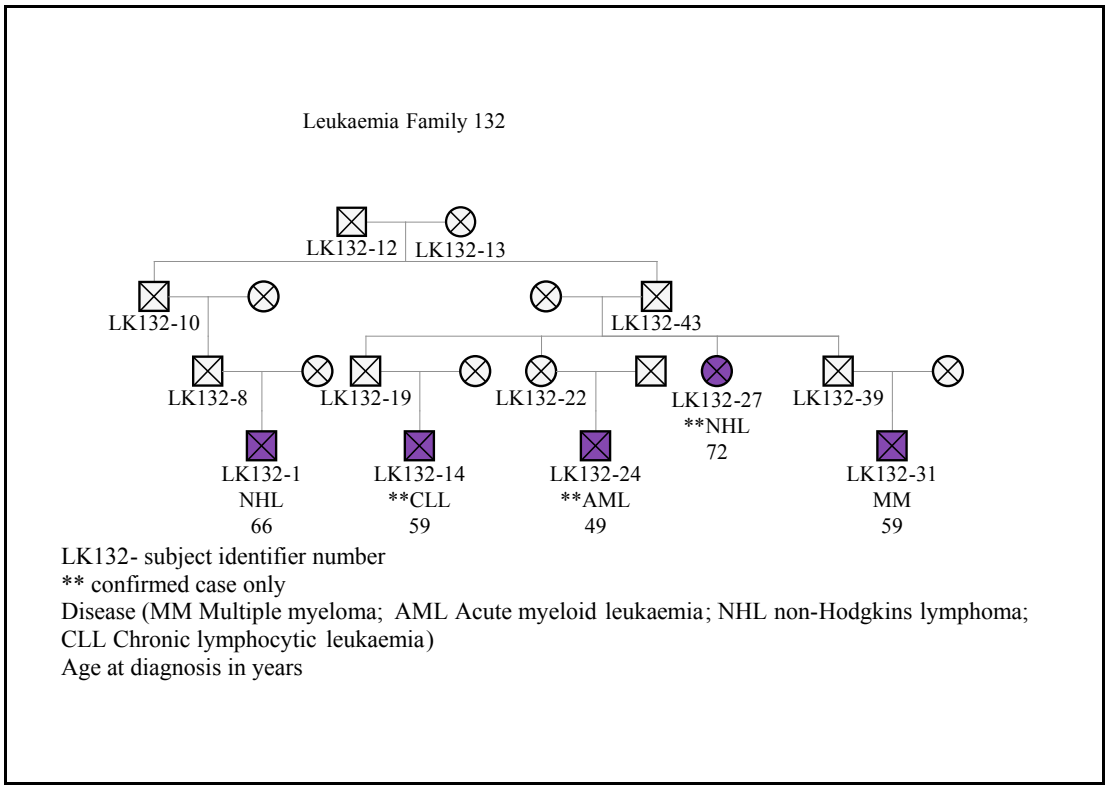


Table 3.3.9.1: The disease characteristics of LK132

ID	Gender	DOB	Date of Diagnosis	Date of Death	Diagnosis*	Age at diagnosis
001	M	13/07/1936	01/1/2002	24/03/2004	NHL	66
014	M	08/06/1925	01/06/1984	24/07/2004	**CLL	59
024	M	10/10/1930	01/08/1979	19/09/1979	**AML	49
027	F	10/11/1910	01/01/1982	10/06/1986	**NHL	72
031	M	19/10/1939	07/04/1998	15/07/1999	MM	59

*are defined in the abbreviation section

** confirmed cases

3.3.10. Leukaemia Family 153 (LK153)

There are 7 cases in this family. An abbreviated pedigree is shown in Figure 3.3.10.1. The disease characteristics are given in Table 3.3.10.1. Only three cases could be reconfirmed in this family. Two cases were from the original study(33, 112, 113). The male to female ratio was 5 to 2. The median age at diagnosis was 57 years (range 34 to 81 years). This family has one sib-pair affected, both of whom are still alive.

Figure 3.3.10.1 Pedigree of LK153

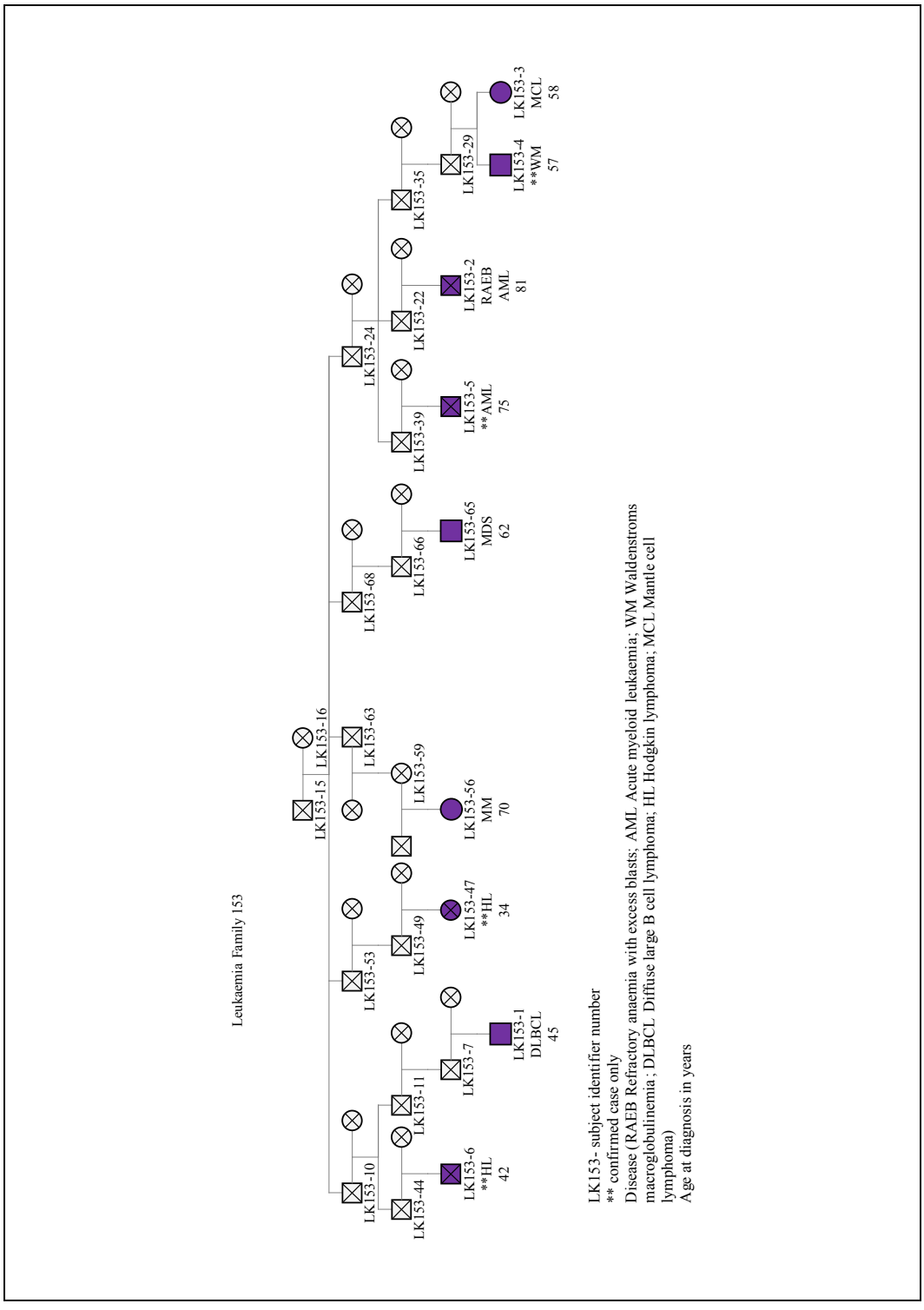


Table 3.3.10.1: The disease characteristics of LK153

ID	Gender	DOB	Date of Diagnosis	Date of Death	Diagnosis*	Age at Diagnosis
1	M	30/09/1956	01/04/2002	Alive	DLBCL	45
2	M	04/01/1924	01/02/2005	30/10/2005	RAEB AML	81
3	F	18/06/1941	01/09/1999	Alive	MCL	58
4	M	12/08/1943	2000	Alive	**WM	57
5	M	03/06/1906	01/05/1981	06/08/1981	**AML NOS	75
6	M	06/07/1909	01/02/1951	02/02/1987	**HL	42
47	F	30/08/1920	01/01/1955	16/06/1955	**HL	34

*are defined in the abbreviation section

** confirmed cases

3.3.11. Leukaemia Family 537 (LK537)

This family consists of a brother and sister who both have CLL (Figure 3.3.11.1). There is no history of consanguinity in this family. Both cases were reconfirmed. The median age at diagnosis was 61.5 years (range 57-66 years).

Figure 3.3.11.1 Pedigree of LK537

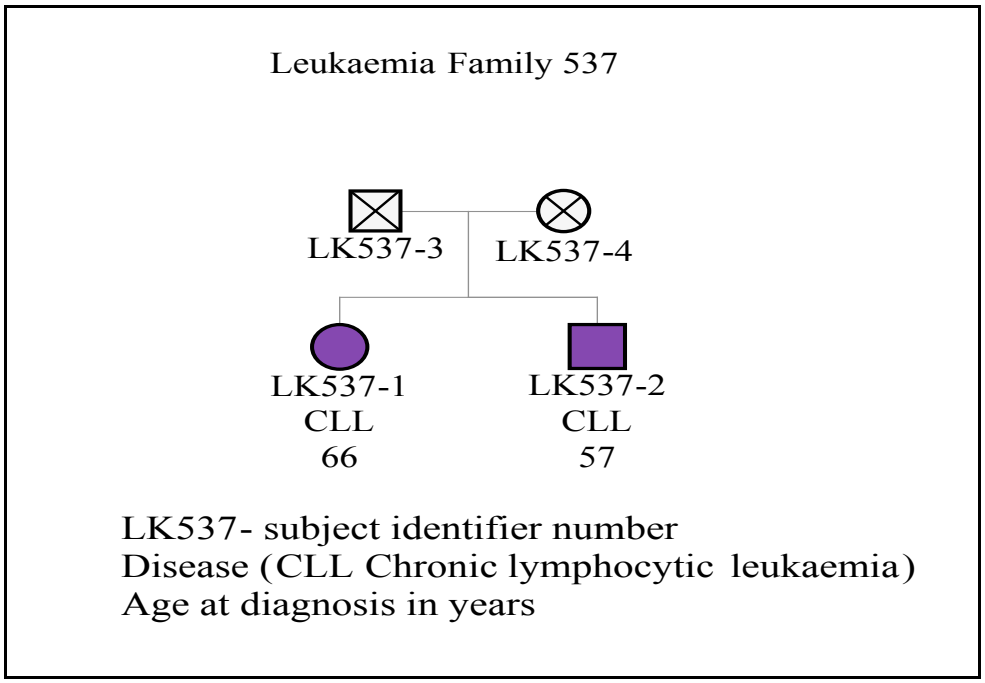


Table 3.3.11.1: The disease characteristics of LK537

ID	Gender	DOB	Date of Diagnosis	Date of Death	Diagnosis*	Age at diagnosis
2	M	28/02/1928	01/11/1985	Alive	CLL	57
1	F	24/06/1937	01/11/2003	Alive	CLL	66

*are defined in the abbreviation section

3.3.12. Leukaemia Family 836 (LK836)

There are 5 cases in this family. A pedigree for this family is given in Figure 3.2.12.1. The disease characteristics are given in Table 3.3.12.1. Three cases in this family could be reconfirmed. Two cases were from the original study(33, 112, 113). All the cases in this family were male and had lymphoid malignancies. The median age at diagnosis was 54 years (range 2 to 70 years). It is interesting to note that 3 of the affected individuals had first cousins as parents.

Figure 3.3.12.1 Pedigree of LK836

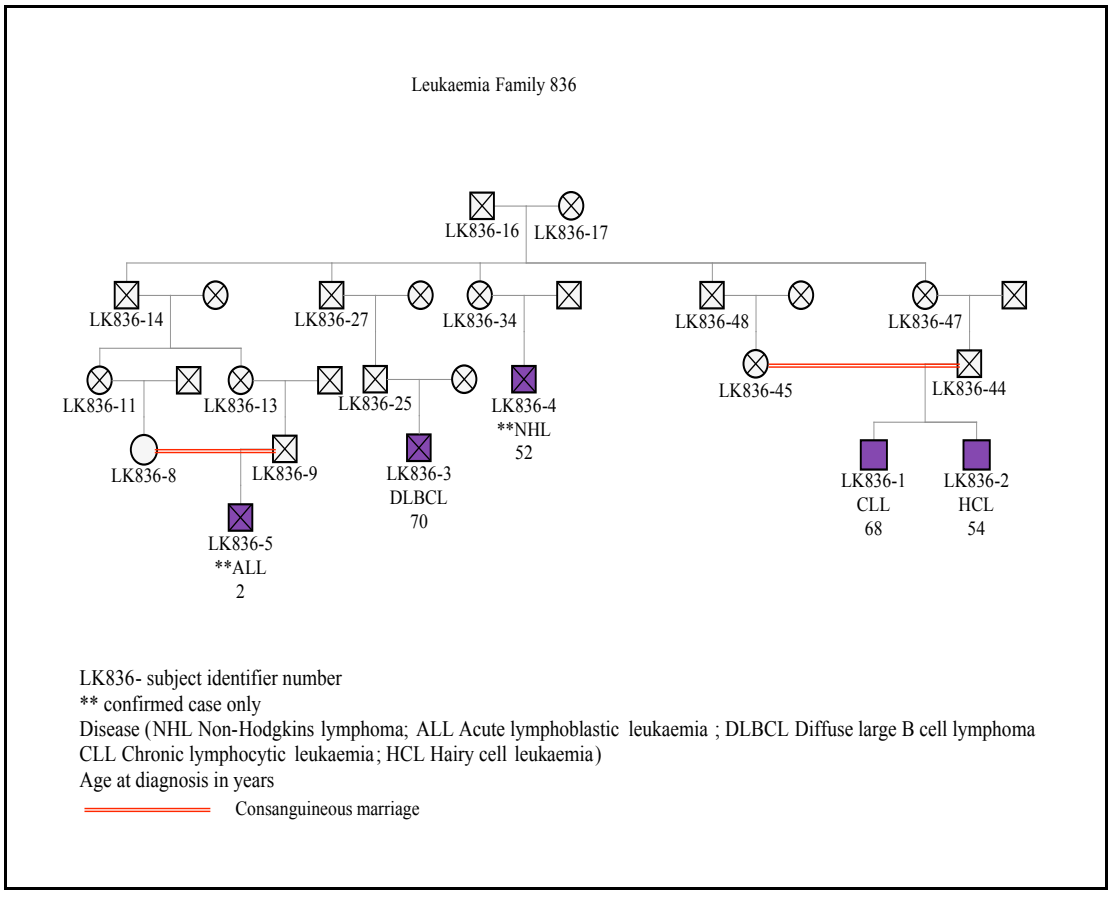


Table 3.3.12.1: The disease characteristics of LK836

ID	Gender	DOB	Date of Diagnosis	Date of Death	Diagnosis*	Age at diagnosis
001	M	12/04/1930	01/05/1995	Alive	CLL	65
002	M	29/08/1933	01/01/1987	Alive	HCL	54
003	M	01/02/1930	01/05/1997	21/11/2000	DLBCL	67
005	M	00/00/1955	1957	08/06/1957	**ALL	2
004	M	07/10/1905	01/02/1957	11/03/1957	**NHL	52

*are defined in the abbreviation section

** confirmed cases

3.3.13. Leukaemia Family 2042 (LK2042)

There are 27 cases in this family. A pedigree is given in Figure 3.3.13.1. The disease characteristics are given in Table 3.3.13.1. Twenty-three cases could be reconfirmed. Only 5 cases were from the original study(33, 112, 113). This is the largest family with 27 affected; the male to female ratio was 21 to 6. The median age at diagnosis was 55 years (range 2 to 86 years). This family contains a range of both myeloid and lymphoid malignancies. One unusual clinical feature is a 40 year old male who was diagnosed with CLL, and then developed CNS involvement 6 years after diagnosis. This family also has four cases of childhood leukaemia; two with B lymphoblastic leukaemia and two with T lymphoblastic leukaemia. The median age at diagnosis for these four childhood cases was 8.5 (range 2-13 years of age).

Figure 3.3.13.1 Pedigree of LK2042

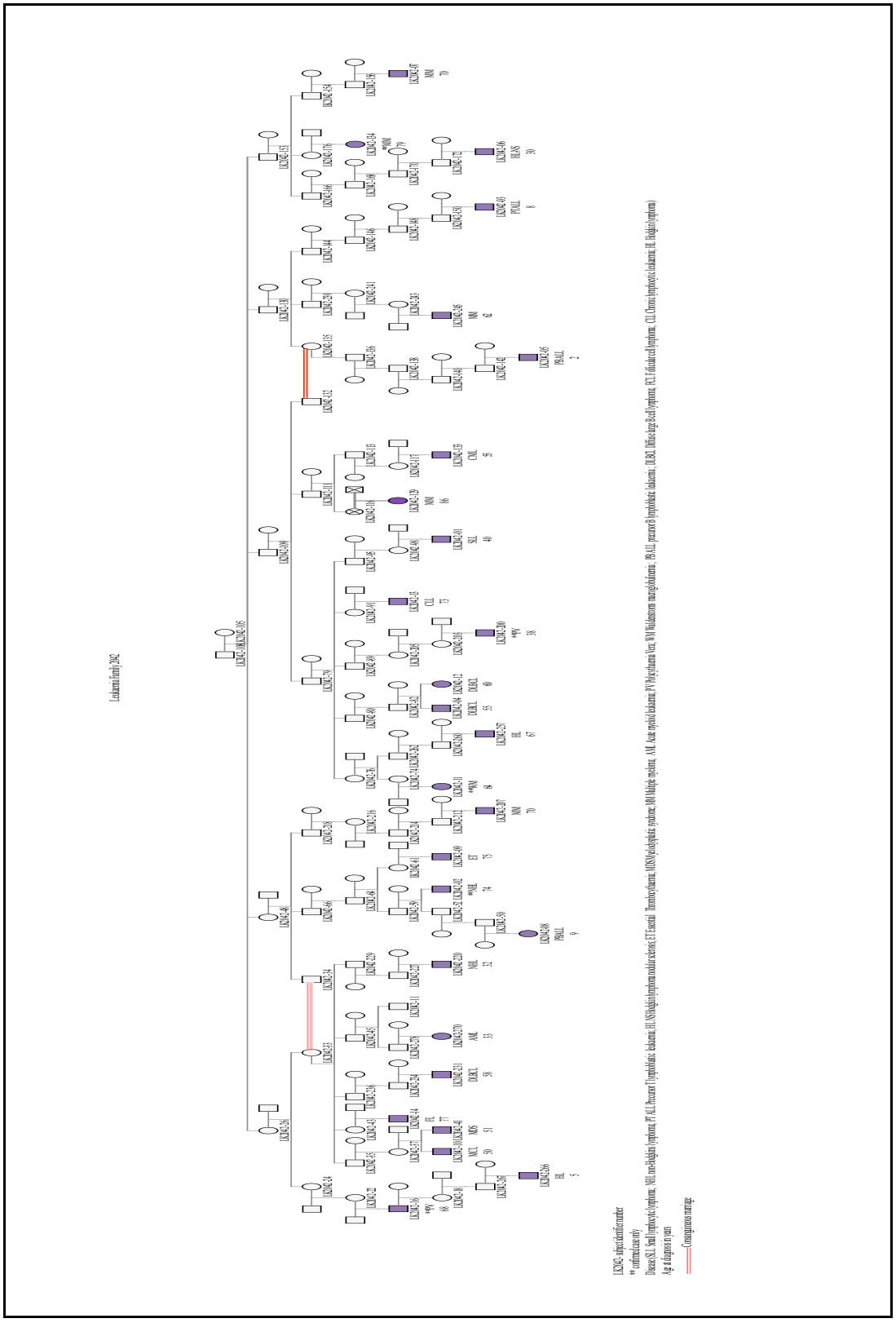


Table 3.3.13.1: The patient characteristic of LK2042.

ID	Gender	DOB	Date of Diagnosis	Date Of Death	Diagnosis*	Age at diagnosis
1	M	24/08/1943	01/04/1983	31/01/1993	SLL	40
2	M	11/04/1918	01/02/1992	12/03/1992	**NHL	74
3	M	27/05/1981	04/05/1989	Alive	PT ALL	8
4	M	8 /07/1927	01/04/1982	21/02/1988	NHL	55
5	M	3 /12/1991	01/03/1994	Alive	PBALL	2
6	M	24/02/1970	01/02/2000	Alive	HL NS	30
7	M	23/03/1930	01/09/2000	11/05/2003	MM	70
8	F	30/07/1970	01/09/1979	27/04/1980	PT ALL	9
9	M	11/10/1924	01/04/1999	31/07/2002	ET	75
10	M	18/03/1944	01/01/1994	2 /09/1994	MCL	50
11	F	20/05/1916	01/10/1980	11/10/1980	WM	64
12	F	29/08/1942	01/03/2002	25/02/2004	DLBCL	60
13	M	30/07/1916	01/02/1989	21/03/1989	CLL	73
14	M	13/06/1899	01/01/1976	10/09/1976	FCL	77
16	M	21/12/1911	01/01/1980	10/01/1981	PV**	68
41	M	03/06/1952	01/09/2003	21/10/2008	MDS	51
129	F	11/02/1919	12/10/2004	Alive	MM	86
133	M	19/12/1932	01/04/1988	20/04/1988	CML	55
134	F	19/03/1898	01/01/1978	27/02/1981	MM**	79
200	M	27/08/1946	01/04/1978	28/12/1995	PV**	32

207	M	7 /12/1978	01/12/1991	19/06/1994	PBALL	13
220	M	03/09/1951	01/11/1983	27/04/1984	NHL	32
231	M	06/07/1941	01/08/1999	24/05/2008	DLBCL	58
245	M	26/02/1940	01/08/2002	29/01/2006	MM	62
257	M	30/12/1939	01/10/2006	Alive	HL	67
266	M	25/07/1995	01/07/2000	Alive	HL	5
270	F	1943	1976	02/08/1976	AML	33

*are defined in the abbreviation section

** confirmed cases

3.4. *Current Tasmanian patients with a family history*

In June 2006, approximately 560 subjects were registered with the TCR as having a confirmed HM and were alive. A total of 400 subjects from this list were traced, current addresses found, and permission obtained from their treating doctor to contact them. This was done to try to identify new families with HMs in Tasmania. One hundred subjects responded to the letter, but only 76 gave information about their family history. Of the 76 who provided information about their relatives with a HM, 15 reported having affected relatives (19.7%; number of affected relatives 1-3). Table 3.4.1 outlines how many were aware of a family history of a HM. There were no cases of parent-offspring affected pairs.

Table 3.4.1: Shows the characteristics of the current patients who responded to our survey from the TCR and their affected relatives.

Family number	DOB	Diagnosis*	Relatives affected
530	04/0719/41	FL	0
558	27/06/1934	CLL	0
560	02/11/1951	NHL	1
589	27/08/1964	PBALL	0
609	16/03/1903	HD	0
612	08/01/1934	NHL	0
614	24/07/1952	HD	0
628	01/12/1930	DLBCL	0
627	18/02/1925	NHL	1

633	02/12/1938	NHL	0
647	24/10/1927	CLL	2
658	31/01/1927	CLL	0
660	13/05/1933	WM	1
673	12/10/1953	NHL	0
692	03/12/1948	CLL	2
718	14/12/1942	MDS	0
732	29/07/1950	TNHL	0
737	20/11/1945	HCL	1
739	14/05/1919	CLL	0
740	12/11/1953	APML	0
741	12/09/1923	MM	0
823	06/03/1951	DLBCL	0
853	25/9/1942	NHL	0
855	11/12/1935	NHL	0
908	08/12/1919	MM	0
933	20/08/1936	HCL	0
946	18/01/1928	CML	0
948	14/03/1957	CML	1
954	30/03/1925	ML	0
958	15/10/1952	HCL	1
1033	16/10/1934	NHL	0
511	20/03/1932	HD	0

512	30/12/1941	NHL	0
586	04/05/1932	MM	0
569	04/07/1951	DLBCL	0
585	06/07/1939	SLL	0
595	24/10/1933	CLL	0
589	05/02/1936	NHL	0
600	10/04/1943	MM	0
635	25/10/1934	NHL	0
650	19/11/1935	CLL	0
664	17/04/1957	FCCL	0
684	13/03/1926	MPD	0
689	16/07/1926	NHL	1
715	25/04/1933	MM	0
717	25/07/1940	NHL	0
760	31/10/1936	NHL	0
932	12/09/1942	MM	0
952	13/09/1959	ALL	0
981	04/12/1938	NHL	0
1037	20/09/1935	NHL	0
502	30/12/1963	MCL	0
508	10/05/1982	AML M5	0
509	28/07/1973	PRV	0
546	09/08/1949	MM	1

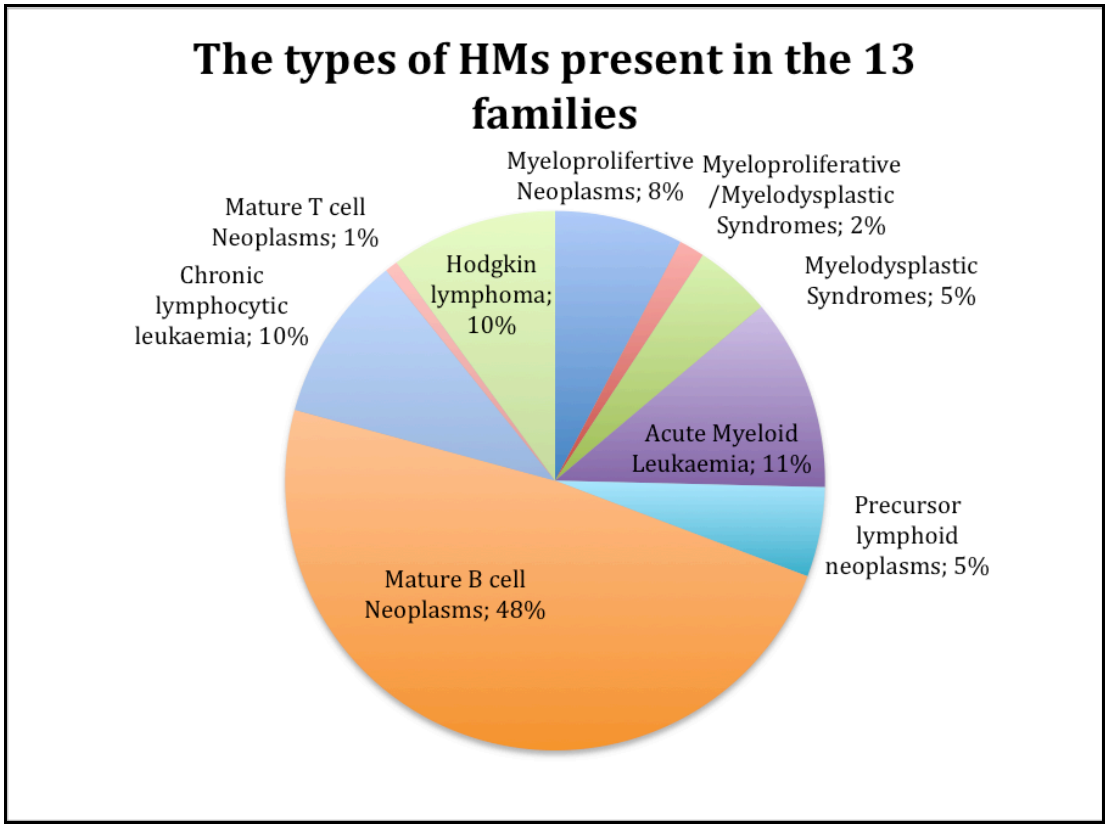
561	18/01/1961	AML	1
613	29/03/1930	NHL	2
625	31/10/1946	MGUS	2
629	31/03/1948	AML	1
634	21/07/1969	HD	0
661	01/01/1932	FCCL	0
672	08/07/1977	PBALL	3
693	24/10/1929	MPD	0
756	21/07/1932	NHL	0
811	08/03/1931	DLBCL	0
818	01/10/1934	DLBCL	0
838	06/11/1934	AML	0
848	10/04/1938	NHL	0
849	14/11/1940	NHL	0
907	24/07/1960	MM	0
956	14/07/1951	NHL	0
984	19/04/1985	HD	0
1035	18/10/1924	NHL	0
1039	14/08/1959	AML	0
1040	17/11/1947	NHL	0
1045	21/07/1971	NHL	0

*are defined in the abbreviation section

3.5. *Results: Summary of the familial dataset*

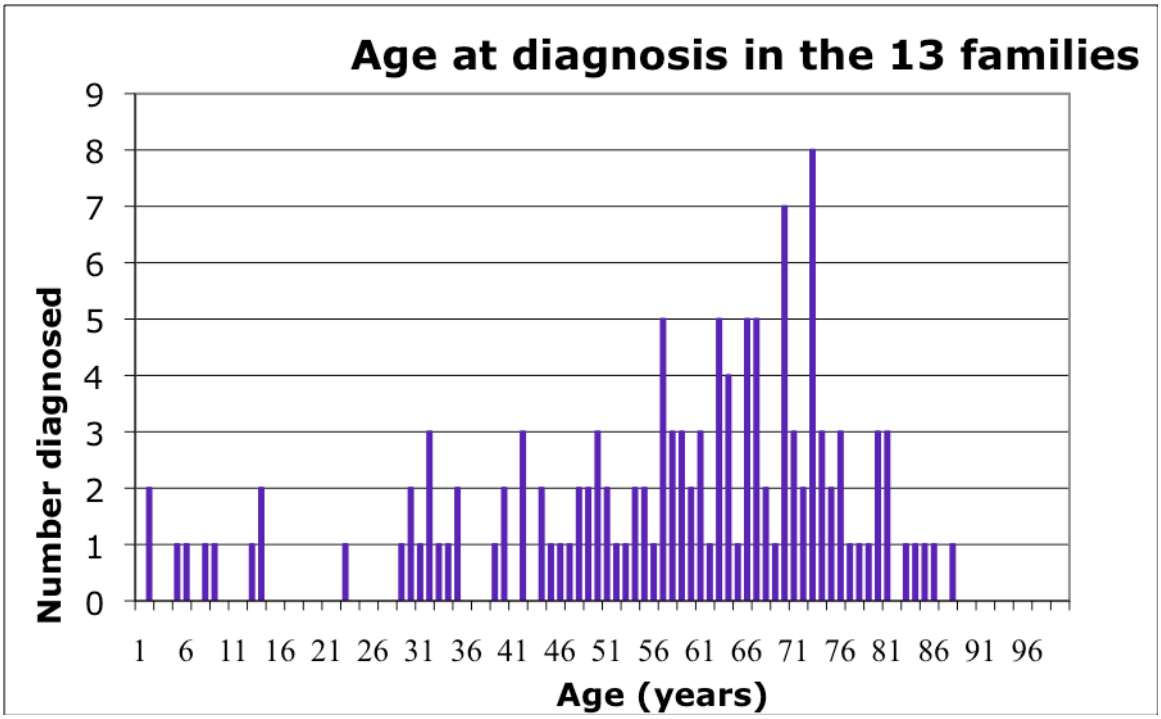
In these 13 families there are 130 affected individuals with a male to female ratio of 95 to 35. A total of 69 additional cases have occurred in these families since the original Tasmanian study(33, 112, 113). Reconfirmation of diagnosis was sought in all cases. A total of 105 could be reconfirmed and classified according to the current haematological classification system(1). The 25 that could not be reconfirmed consisted of 4 diagnosed outside of Tasmania and a further 21 who were in the original studies(33, 112, 113) but whose pathology records are now no longer available. Figure 3.5.1 shows the types of HMs present in these families according to the major subgroups in the WHO classification system(1). Pathology samples have been located for 90% of the affected cases.

Figure 3.5.1 Types of HMs in the 13 families.



The age at diagnosis and the type of HM affecting those 25 persons was recorded in the original study(33, 112, 113). The median age at diagnosis for all 130 affected individuals was 63 years (2-86 years). Figure 3.5.2 shows the age at diagnosis of all cases. The figure shows a tri-modal age distribution, with a peak in childhood (age <16) another peak between 16-36 years but with the majority of cases presenting in the sixth and seventh decades.

Figure 3.5.2 Age at diagnosis for all cases (n=130).

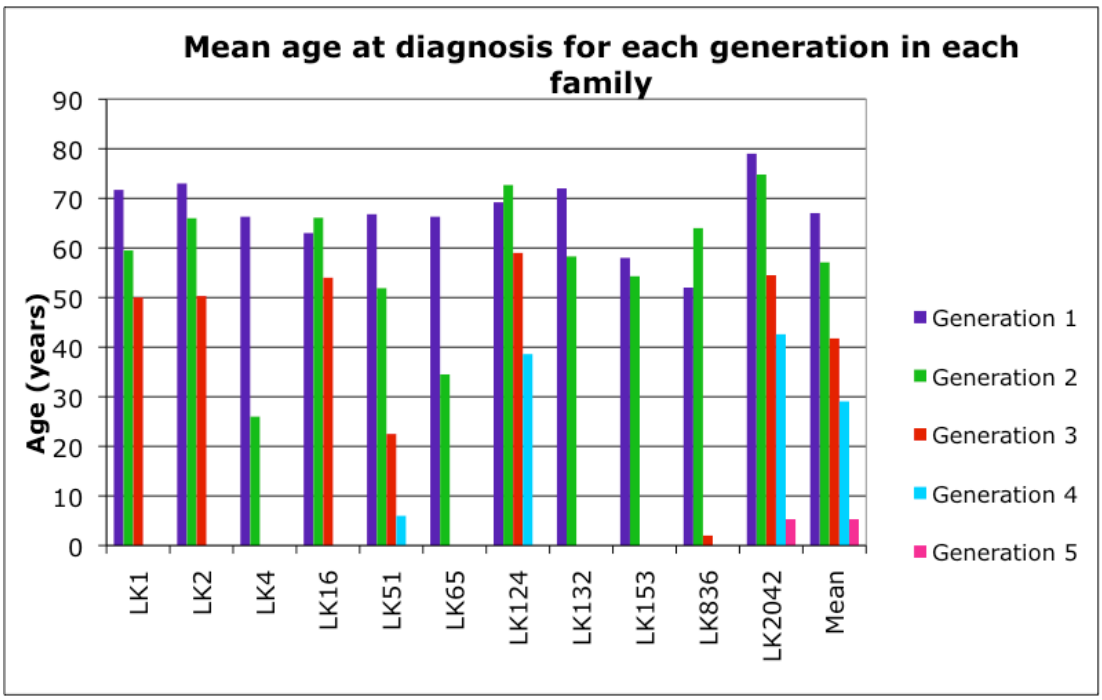


Eleven families out of the thirteen studied had multiple generations affected with a HM. The age at diagnosis was assessed for each generation (Table 3.5.1 and Figure 3.5.3). In nine of the families, each affected generation after the first exhibits a reduction in age at diagnosis. Families LK16 and LK836 do not show this phenomenon.

Table 3.5.1: Mean age at diagnosis [years (number of subjects)] for each generation.

	G1	G2	G3	G4	G5
LK1	71.7 (4)	59.5 (4)	50.0 (1)		
LK2	73 (3)	66.0 (6)	50.3 (3)		
LK4	66.3 (1)	26.0 (2)			
LK16	63 (1)	66.1 (9)	54 (2)		
LK51	66.8 (5)	51.9 (7)	22.5 (2)	6.0 (1)	
LK65	66.3 (4)	34.5 (4)			
LK124	69.2 (7)	72.7 (7)	59.0 (1)	38.6 (3)	
LK132	72 (1)	58.3 (4)			
LK153	58.0 (4)	54.3 (3)			
LK836	52 (1)	64 (3)	2 (1)		
LK2042	86.0 (1)	74.8 (5)	54.5 (13)	42.6 (5)	5.3 (3)

Figure 3.5.3: Graph of the mean age at diagnosis (years) for each generation for all the families that have at least two generations affected.



In the nine families where there was a reduction of the age at diagnosis, there were five direct parent-offspring cases in these nine families. The age at diagnosis for these cases is in Table 3.5.2. The mean age at diagnosis for the parents was 63.8 years and for their offspring 44.8 years, this was a statistically significant decrease in the age at diagnosis by 19 years ($p=0.0087$). A log rank test for trend of the age at diagnosis by generation was statistically significant (see Figure 3.5.4; $p<0.0001$). The types of HMs in each generation was assessed (Figure 3.5.5), however no bias of subtypes such as ALL that have a younger age of onset was found, in particular MBCNs was diagnosed across four generations and ALL was also diagnosed across four generations. The age at diagnosis was assessed in one subtype of HM, MBCN. This was the largest type of diagnosis represented in these families and the subtypes of MBCN have a similar age of onset. Figure 3.5.6 demonstrates the age at diagnosis

for all cases of MBCN in families with at least two generations affected with this type of HM. This also shows a significant decrease in the age at diagnosis for each subsequent generation (P=0.0007).

Table 3.5.2: Age at diagnosis (years) for direct parent-offspring affected pairs.

Family	**LK1**	**LK1**		**LK2**			**LK4**		**LK51**	**Mean ± SD**
Age of parent	72	73		60			73		51	**63.8 ± 9.1**
Age of offspring	50	61	47	66	39	46	29	23	42	**44.8 ± 13.7**
P value										**0.0087**

Figure 3.5.4: Test for trend of the age at diagnosis (years) for each generation

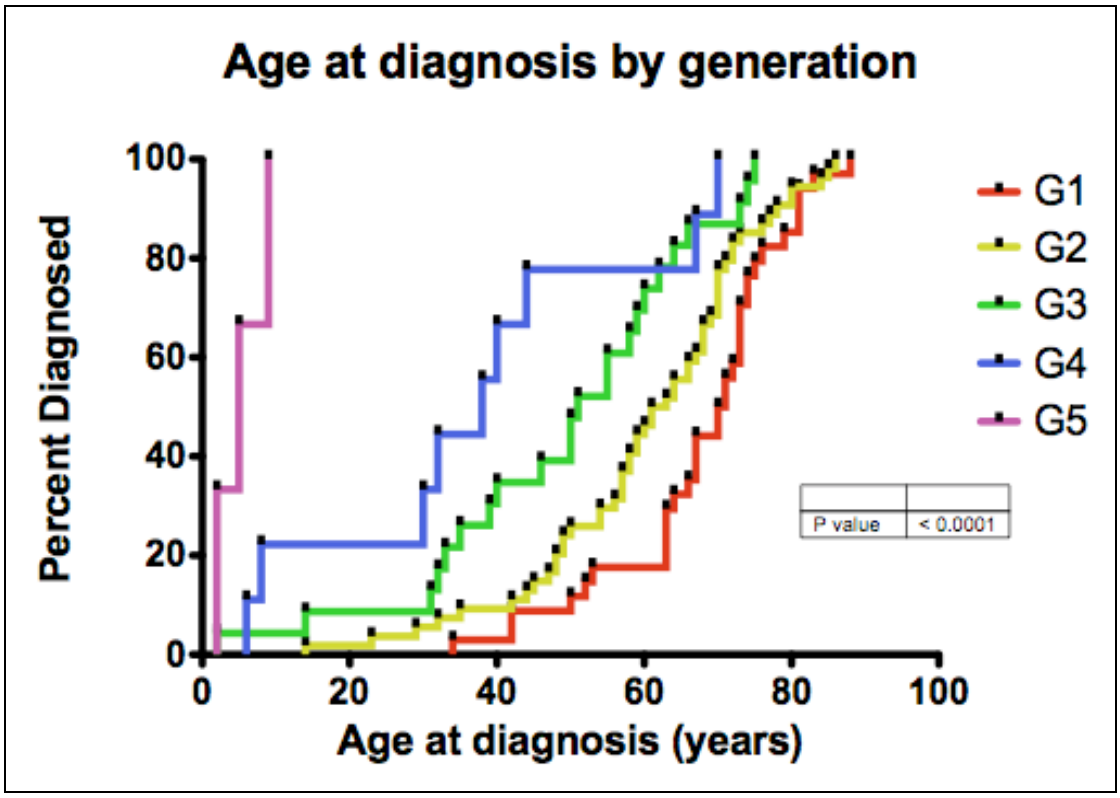


Figure 3.5.5: Distribution of the type of haematological malignancy in each generation.

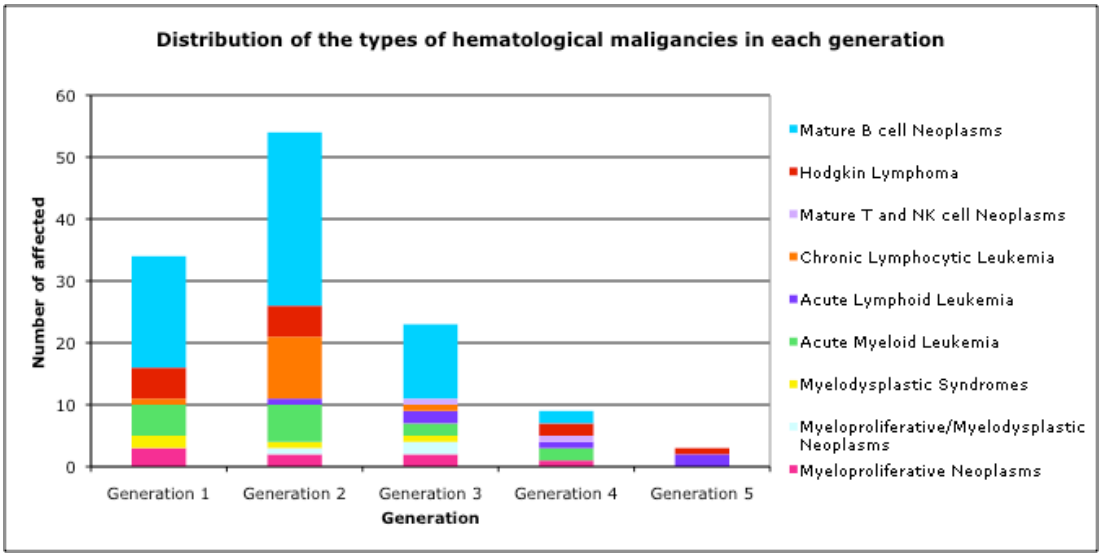
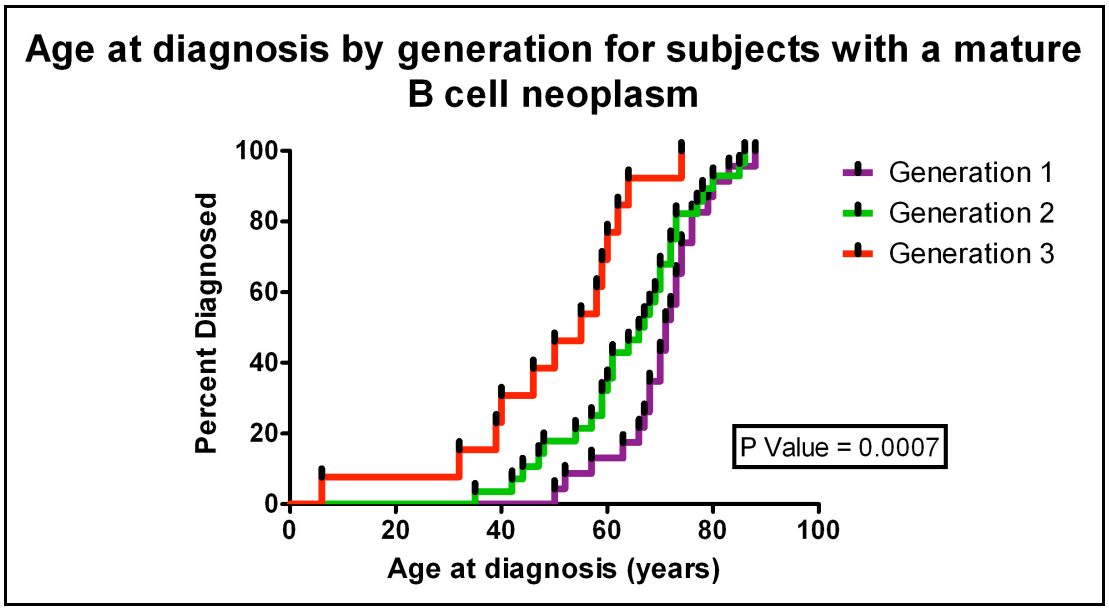


Figure 3.5.6: Age at diagnosis for each generation for subjects with a Mature B cell neoplasm



3.6. *Discussion and Conclusions*

The 13 Tasmanian families in this study are a unique research resource because of the number of cases of HM confirmed diagnosis and recording of their genealogy. In nine of the families there is a reduction in the age at diagnosis for each subsequent generation. This is confirmed in direct parent-offspring affected cases with a mean reduction of 19 years in the age at diagnosis between the generations. A decrease in age at diagnosis for each subsequent generation is often reported in familial disorders and is referred to as anticipation. It is said to be due to an autosomal dominant trait(7, 22). Anticipation in CLL families with a reduction of 15 years in the age at diagnosis has been reported in the literature(7). It has also been observed in familial AML families, with the mean age at diagnosis in the first generation being 57 years, in the second generation 32 years, and in the third generation, 13 years(10). Anticipation

has also been observed in familial lymphoma families(23), in families with plasma cell dyscrasia(5) and in familial Hodgkin lymphoma(20). In fact, over 140 families with familial HMs which show anticipation have been documented(17). As different subtypes of HMs have different median age of onsets, the types of HM found in each generation of these families was assessed (Figure 3.5.5.), this showed no skewing of the subtypes of HMs in the younger generations. However, to confirm the anticipation, the analysis of the age at diagnosis was repeated for all cases of mature B cell neoplasm (including CLL), as they have a similar age of onset (Figure 3.5.6) (therefore, addressing the argument of young onset diseases only occurring in the younger generations) and there was still a significant decrease in the age at diagnosis for each subsequent generation ($p=0.0007$)(162).

Anticipation is reported to be consistent with an autosomal dominant mutation**.** In instances where anticipation occurring in familial disorders has been reported, the mutation if known has been due to a triplet repeat mutation(139). However, the mutation in HMs has not been identified, although three genetic regions have been suggested to be involved in HMs; a locus on 21q22.1-22.2, a locus on 11q23.3 and a locus on 16q22 (10). It is interesting to note that the two families from the 13 Tasmanian families that do not show a consistent decrease in age at diagnosis, were those families in which there was consanguinity involved (LK16 and LK836). Both of these families were selected because of a cluster of affected siblings whose parents were first cousins, suggesting that a different genetic mechanism (recessive mutation) is involved in these families (163).

Consanguinity is a feature of some of the families in this study, in particular families LK16, LK836 and LK2. Consanguinity in these families consists of first cousin

marriages. There were six sets of consanguineous marriages, which resulted in 27 children, of whom 11 were affected with a HM. A Japanese study from the 1970s found a significant increase in the number of consanguineous marriages in families with familial AML compared to the general population(12). It is known that consanguineous marriage leads to an increased prevalence of infants with recessive disorders(164). Therefore, a recessive mutation may be implicated in these families. It should be noted the LK2 family that does have consanguineous marriages also shows anticipation. This raises the possibility that the genetic predisposition in this family is not due to simple Mendel's genetics, but rather due to a more complex genetic predisposition.

The clinical features of the HMs in these 13 families studied were also reviewed. Three cases of primary CNS lymphoma were noted in the LK124 family. This is a rare site for lymphoma to present(165). The LK16 family consisted of 5 affected siblings with a HM, 4 with CLL and one with DLBCL. Houlston *et al.* 2003 has reported in the literature that "a family with three siblings affected by chance would be expected to occur once every 100 years in the UK"(11), thus making the LK16 family in Tasmania with five affected siblings, where the population is significantly smaller than in the UK, is extremely rare. In fact, in these 13 Tasmanian families, there is 1 set of 5 siblings affected by a HM, 1 set of 3 siblings affected and 11 sets of affected sibling pairs.

In the LK2042 family, CNS involvement is noted in a case of CLL, which according to the literature, is rare, as there are only 24 reported cases of this occurring(166). There was also an unexpectedly low number of CLL cases in these Tasmanian

families, as CLL is reported to be the most common leukaemia in western countries(6). In these families there were more cases of ALL or AML than CLL.

The current Tasmanian survey of living affected people registered by the TCR revealed no new large families with HMs; however, approximately 20% of those who responded, had a positive family history of a HM. This is compared to 8% found by Chang *et al*. 2005(3), in the Swedish population. The higher positive rate in the Tasmanian population may reflect the low response rate to the survey (25%) and the response rate bias of people with a positive history being more inclined to respond to the survey than someone who has a negative family history.

It is clearly recognised in the literature that there is a genetic basis for the familial risk of HMs(3-17, 19, 22-24, 130, 167, 168). However, although genome-wide studies have identified several genetic loci, these only account for approximately 3% of the inheritable risk(149). Additional loci for CLL and ALL have been recently published, but the SNPs identified so far for CLL are still reported to account for only 10% of the inheritable risk(144). There is evidence provided by other studies(169) of familial disease (Parkinsons disease) where application of high-density genotyping and next generation sequencing to the study of familial disease will enable the identification of those rare alleles. These studies are likely to complement genome wide association studies, which are not designed to detect the rare alleles. Further study of these 13 Tasmanian families will be able to further explain the heritable risk in HMs.

4. Chapter 4: The LK16 family with 5 affected siblings from a consanguineous marriage

4.1. Introduction

Large families with multiple cases of HMs are rare(19). The vast majority of families that are reported in the literature have familial CLL(6, 8, 11, 15, 18, 170, 171). The largest reported pedigree was an Australian family(136). This family reported by Fuller *et al.* 2008, consists of 11 subjects with CLL over 3 generations, with several members of the fourth generation having monoclonal B cell lymphocytosis (MBL)(136). Another large CLL pedigree reported by Lynch *et al*. 2002, consists of a father and 4 sons affected with CLL(170), and more recently reported a third generation with CLL(25). This is the only family in the literature where a germline mutation has been identified and this mutation results in decreased expression of *DAPK1*(25).

A genome-wide linkage study of 206 families with CLL (18) investigated the familial cause of CLL. This linkage study failed to find a definitive genetic cause, however it found that the region showing maximum nonparametric linkage was to chromosome 2q21.2; and two other regions also showed significant linkage (6p22.1 and 18q21.1)(18). The region on 6p22.1 was a recessively acting locus in contrast to a dominantly acting locus on 18q21.1(18).

There are many types of SNP arrays available for use, such as oligonucleotide arrays (Agilent), Affymetrix SNP arrays and Illumina SNP arrays(172). The Affymetrix GeneChip® 250k Nsp SNP array was selected, as it was the most cost effective high

density SNP array available for use. It contains over 262,000 SNPs and uses established technology from the GeneChip® Mapping 10K array.

The LK16 family was chosen to study first as there were 5 siblings affected with a HM whose parents were first cousins. Their mother was also affected with a HM and died as a result of CML in 1943 and six other members of this family are also affected with a HM. A pedigree is given in chapter 3.3.4. It is stated in the literature that "a family with three affected siblings would be expected to occur by chance approximately once every 100 years in the UK" (173), thus this represents a very rare family, even considering the purported selection bias that exists in the identification of families featuring a high aggregation of cases. There are a total of 12 subjects affected in this family across 3 generations.

It is known that the offspring of consanguineous relationships are at an increased risk of recessive disorders(163, 164), as they inherit the same mutation from each parent. It is hypothesized in this family that the five siblings are homozygous for a region of the genome where the mutation is located. Therefore offspring of the affected siblings will be heterozygous for this region.

4.2. Materials and Methods

4.2.1. Sample Collection

Living affected members and first-degree relatives of deceased affected members were contacted and asked to consent to participate in this study and to provide a blood or saliva sample for DNA extraction as described in Chapter 2.1. DNA from LK16-1, LK16-106, LK16-4, LK16-5, LK16-6, LK16-7, LK16-8, LK16-11, LK16-

15, LK16-16 and LK16-3 were sent to the Australian Genome Research Facility for genome-wide Affymetrix GeneChip Human Mapping 250K Nsp SNP Array.

4.2.2. Locating Pathology samples

Formalin fixed paraffin embedded tissue (FFPET) was located at Bendigo hospital on LK16-14, who was diagnosed with melanoma from a skin biopsy in July 1978. FFPET from LK16-12 was located from Launceston Pathology and consisted of a skin biopsy and an ethmoid sinus biopsy. FFPET from LK16-9 was located from Launceston General Hospital and consisted of a gastrointestinal biopsy. Bone marrow biopsy samples from LK16-101, LK16-87, and LK16-50 were located from the RHH and peripheral blood from LK16-108 who was diagnosed with multiple myeloma in January 2008. DNA was extracted as per Chapter 2.2.

4.2.3. Statistical analysis

Affymetrix GeneChip® Human Mapping 250k Nsp SNP Array data was analysed by Dr Devindra Perera using Merlin (Gonalo Abecasis) analysing the genotypes of the genome where LK16-1 was homozygous and the children of the other affected siblings were heterozygous (IBD sharing). Haplotype analysis was performed using Merlin (Gonalo Abecasis) and HaploPainter. Pedigrees were drawn with Smartdraw version 5.0.

4.2.4. PCR primers

Primers for microsatellite markers and genes were designed and optimised as per Chapter 2.6. Sequencing of genes was on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems) as described in Chapter 2.8.

4.2.5. Conventional Cytogenetics and FISH

Lithium heparin peripheral blood samples were obtained from LK16-4, LK16-15 and LK16-16. Chromosome preparations were obtained as described in Chapter 2.9. FISH probes for *IGH@*, 13q14, 13q34, CEP(12), *TP53* and *ATM* were obtained from Vysis (Abbott Molecular, Australia) and used as described in Chapter 2.10.

4.2.6. Laser Microdissection of the region of interest

Chromosome preparations from LK16-15, LK16-16 and LK16-4 were obtained. The region of interest on chromosome 15 was dissected (as described in chapter 2.12), DNA isolated (as described in Chapter 2.2) and whole genome amplified using GenomePlex® (Sigma-Aldrich) (as described in Chapter 2.4).

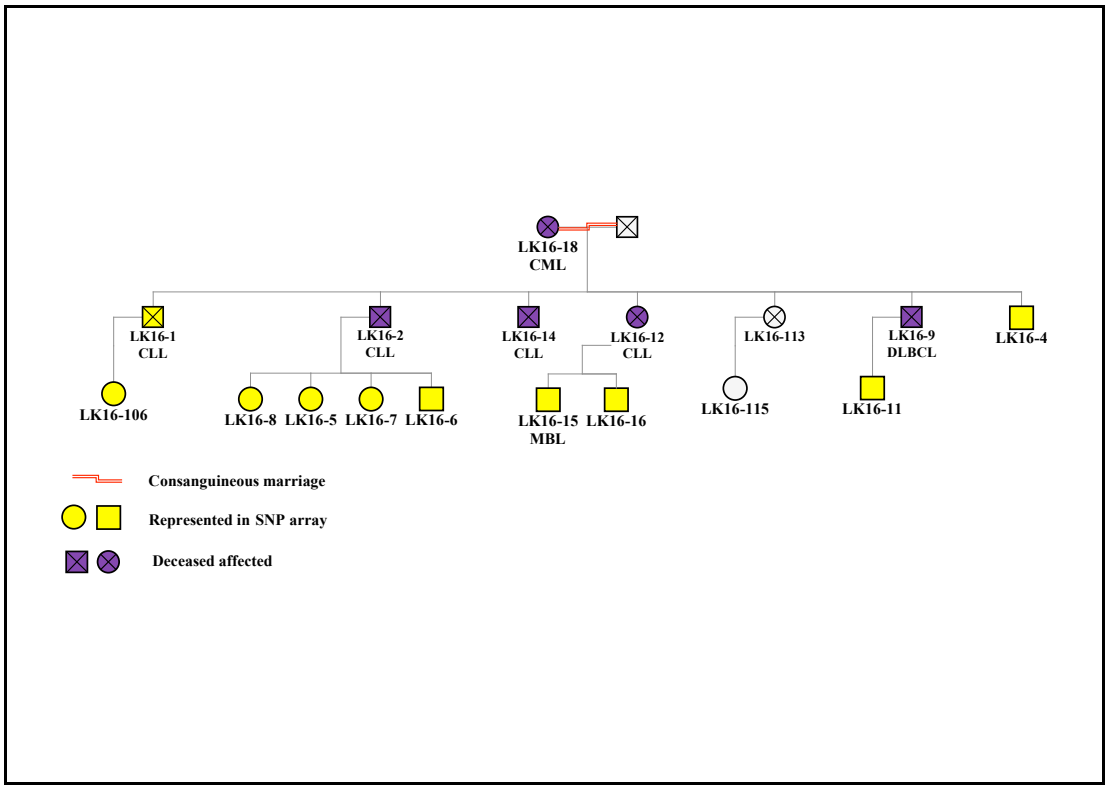
4.3. Results

4.3.1. Statistical analysis of the Affymetrix GeneChip® Human Mapping 250k Nsp SNP Array

At the time of this study only one of the affected siblings (LK16-1) was alive, and he died of his disease shortly after consenting for this study. His child and children of his deceased affected siblings (LK16-9, LK16-12 and LK16-2) and his sole surviving sibling (LK16-4) consented to participate in this study. The other affected sibling LK16-14, had no children. Figure 4.3.1.1 highlights the sub-pedigree structure of those who participated in the study. The genotypes were analysed using a SPLITPED analysis by Merlin (Gonalo Abecasis) analysing for areas of the genome where LK16-1 was homozygous and his nieces and nephews were heterozygous for that

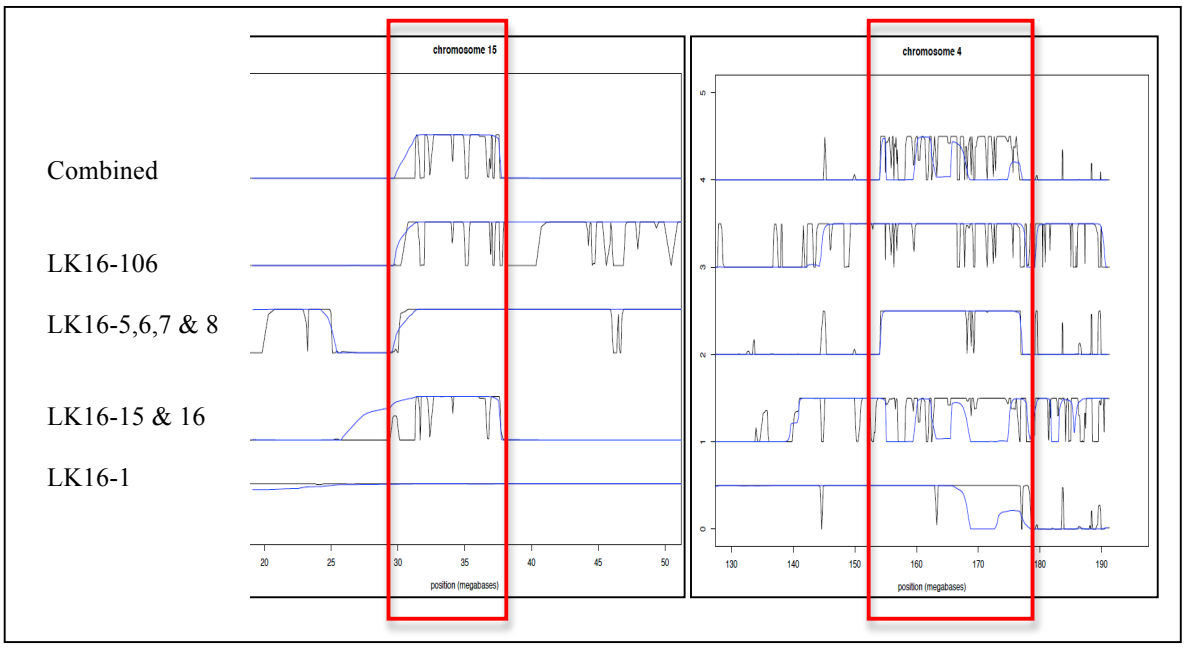
region (IBD). Therefore implying that their parents (LK16-1's siblings) were at least heterozygous for that region.

Figure 4.3.1.1: A pedigree of LK16 family used in the initial analysis.



This analysis identified two regions, one on chromosome 4 and the other on chromosome 15, as shown in Figure 4.3.1.2. DNA extracted from FFPET from the affected siblings was used to confirm which of these two regions of homozygosity in LK16-1 his siblings shared.

Figure 4.3.1.2: Regions of IBD sharing on chromosome 4 and chromosome 15.



The combined regions where LK16-1 is homozygous and nieces and nephews' are heterozygous are the regions within the red box (chromosome 15: 30000000-38000000 and chromosome 4: 150000000-171000000).

4.3.2. Confirming the region of homozygosity in the other affected siblings

4.3.2.1. Microsatellites

Chromosome 4 region

Six microsatellites were chosen across this region (D4S2962, D4S3049, D4S2980, D4S2918, D4S2952, D4S1597). Not all primers would hybridise with the low quality DNA obtained from the FFPET. Table 4.3.2.1 shows the allele sizes for these microsatellites for the siblings. It was concluded that the other siblings are not homozygous for this region on chromosome 4.

Table 4.3.2.1: Microsatellite results for the chromosome 4 region

Microsatellite	Position	LK16-1*	LK16-12	LK16-9	LK16-14	LK16-4
D4S2962	Chr4:150583446-150583787	132/132	132/132	132/132	132/142	132/142
D4S3049	chr4:154993900-154994255	231/231		231/226	231/244	231/221
D4S2980	chr4:160827646-160827989	146/146		146/144	146/148	146/148
D4S2918	chr4:160548592-160548915	160/160		160/162	160/162	160/162
D4S2952	chr4:166675026-166675407	186/186	186/186	186/186	186/188	186/186
D4S1597	chr4:170079690-170080082	274/274		274/274	274/274	274/274

*LK16-1 is homozygous for this region

Chromosome 15 region

Ten microsatellites were chosen across this region (D15S144, D15S995, D15S1040, ACTC, D15S971, D15S118, D15S102, D15S194, D15S1044, D15S214). Table 4.3.2.2 shows the results for this region for the siblings.

Table 4.3.2.2: Microsatellite results for the chromosome 15 region

Microsatellite	Position	LK16-1	LK16-12	LK16-9	LK16-14	LK16-4
D15S144	chr15:31388179-31388405	158/158	158/152	158/158	158/158	158/158
D15S995	chr15:31453771-31454043	192/192		192/192	192/192	192/192
D15S1040	chr15:31913662-31913895	203/203		203/203	203/203	203/203
ACTC	chr15:32870697-32870925	232/232	232/232	232/232	232/232	232/232
D15S971	chr15:33171120-33171439	214/214		214/214	214/214	214/214
D15S118	chr15:34024022-34024377	218/218	218/218	218/218	218/218	218/218
D15S102	chr15:35762875-35763087	210/210	210/210	210/210	210/210	210/210
D15S194	chr15:36136329-36136718	251/251		251/251	251/251	251/251
D15S1044	chr15:37456254-37456576	180/180		180/180	180/180	180/180
D15S214	chr15:38187526-38187868	264/264	264/264	264/264	264/264	264/264

The microsatellite sizes were also assessed on the offspring of the affected siblings (LK16-106, LK16-11, LK16-5, LK16-6, LK16-7, LK16-8, LK16-15 and LK16-16) and they all shared the haplotype in the heterozygous state. This helped to confirm the reliability of the microsatellite results form the DNA extracted from the FFPET. The microsatellite data confirmed that this was a region shared in a homozygous state by the affected siblings and therefore indicated that the chr15: 30000000-38000000 is the region of interest in this family.

4.3.2.2. Affymetrix GeneChip® Human Mapping 250k Nsp SNP Array using DNA from FFPET

DNA was extracted from FFPET from LK16-14, LK16-9 (affected siblings of LK16-1) and LK16-87 (distant relative with CLL). This was sent to the AGRF for the same genome-wide Affymetrix GeneChip Human Mapping 250k Nsp SNP Array. Assessing Mendelian errors using Merlin (Gonalo Abecasis) between offspring and parent pairs enabled the quality of the SNP array results on the FFPET samples to be compared. There were two parent offspring pairs (LK16-1, LK16-106 and LK16-9, LK16-11). There were 782 SNPs in the region of interest on chromosome 15 from the Affymetrix GeneChip® Human Mapping 250k Nsp SNP Array. There was a 12.5% Mendelian error rate between LK16-11 and LK16-9, compared to an error rate of 0.02% between LK16-1 and LK16-106. There was also a very high no-call rate on the samples that were from FFPET (LK16-9 was 26.9%: LK16-14 was 27.9%) compared to 2.5% for the peripheral blood sample from LK16-1. Both the no-call rate and Mendelian error rate were considered too high and therefore the Affymetrix GeneChip® Human Mapping 250k Nsp SNP Array results from the FFPET were not reliable and could not be used.

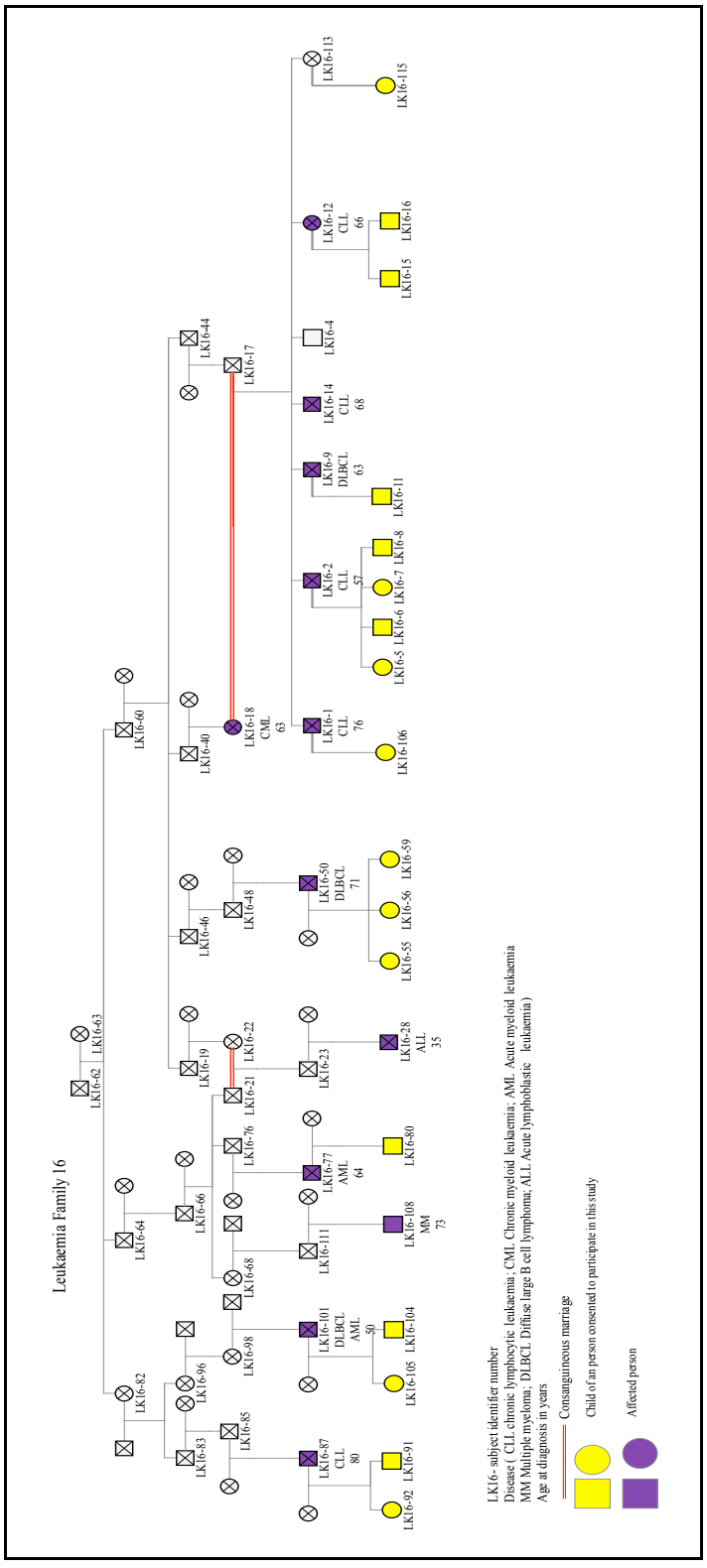
4.3.3. Genes

In the area of interest on chromosome 15, there are 33 genes, many of which have many isoforms. The functions of genes were reviewed in PubMed (http://www.ncbi.nlm.nih.gov/pubmed), however there was no obvious candidate. Several genes that had been implicated in carcinogenesis were sequenced first. The following genes have been sequenced: AVEN, GPR176, MEIS2, PGBD4 and C15orf41. No mutations were found in the coding regions, 5'UTR, 3'UTR or 500bp either side of the gene.

4.3.4. Haplotype in distant affected relatives in the LK16 family

There are a total of 12 people affected with a HM in this family. An abbreviated pedigree of this family is shown in Figure 4.3.4.1. This shows the other affected in this family and who was alive and consented to participate in this study. 18 children of affected deceased members agreed to participate in this study.

Figure 4.3.4.1: Extended pedigree of LK16 including all who have consented to participate in this study.



Microsatellite analysis for the full region of homozygosity was assessed on all available people in this family. The other affected people in this family are LK16-50 (DLBCL), LK16-87 (CLL), LK16-101 (DLBCL and AML), LK16-77 (AML) and LK16-108 (MM) as shown in Figure 4.3.4.1. As the siblings were homozygous for the region on chromosome 15, by default their parents LK16-18 and LK16-17 must also be heterozygous carriers of this region. Thereby implying that their fathers (LK16-40 and LK16-44 respectively, whom are brothers) are also heterozygous carriers and either inherited this from their mother (LK16-61) or father (LK16-60). As the grandfather (LK16-46) of LK16-50 and the great-grandfather (LK16-19) of LK16-28 are also brothers of LK16-40 and LK16-44, there is a 50% chance that these two people (LK16-19 and LK16-46) were also carriers of this haplotype. If they did in fact inherit this haplotype as well, then there is a 12.5% chance that LK16-28 is also a carrier of this haplotype and 25% chance that LK50 is a carrier. From microsatellite and SNP analysis across the larger chromosome 15 region, they did not share this haplotype

4.3.5. Chromosome analysis

Conventional cytogenetic analysis was undertaken in the family to exclude a structural chromosome abnormality segregating in this family, especially at the site of the region of interest on chromosome 15. The chromosome analysis revealed a normal male karyotype for LK16-4 (Figure 4.3.5.1) and LK16-16 (Figure 4.3.5.2). However LK16-15 (Figure 4.3.5.3 normal) showed a translocation between the long arm of chromosome 12 and the long arm of chromosome 14 and a deletion of the long arm of chromosome 13 (Figure 4.3.5.4) in approximately 5% of cells. This was further investigated by FISH with a Vysis IGH@ dual colour break-apart probe. This

confirmed an IgH rearrangement (Figure 4.3.5.5) and deletion of the long arm of chromosome 13 in approximately 5% of cells.

Figure 4.3.5.1: LK16-4 normal male karyogram.

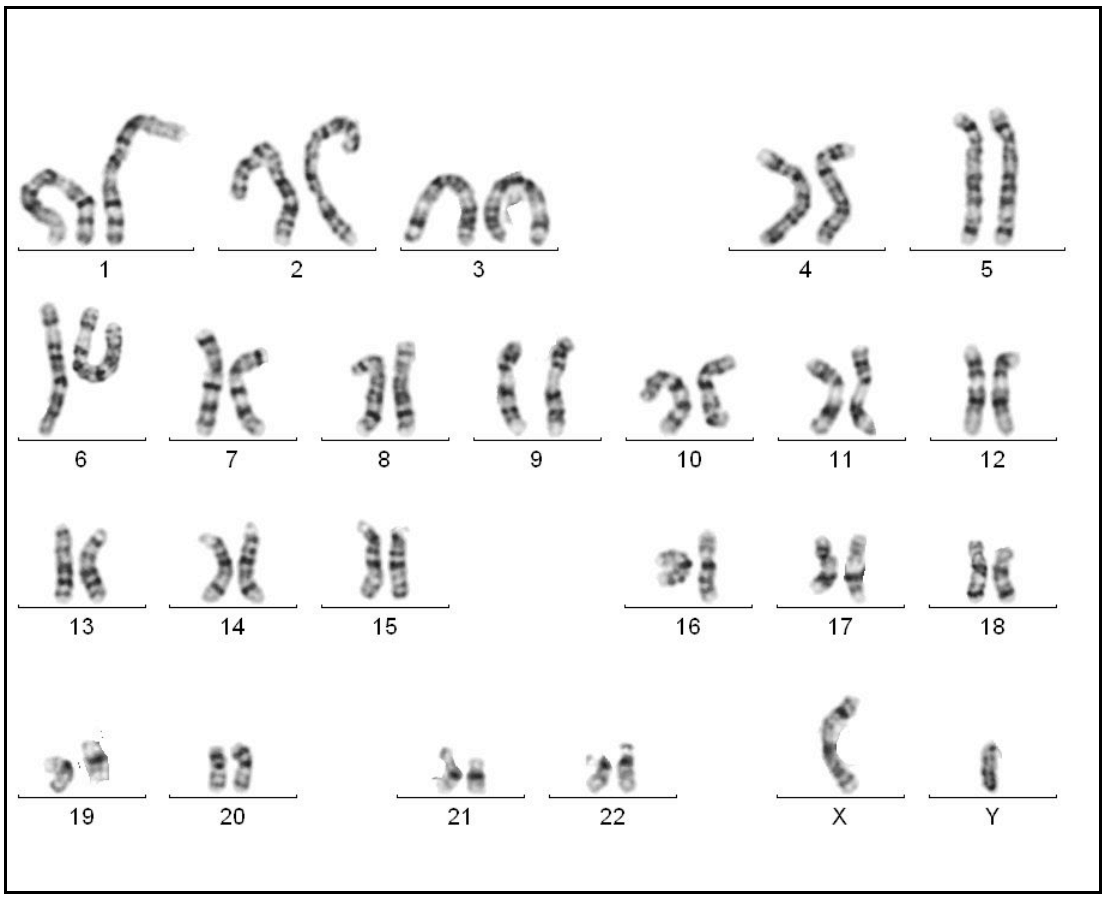


Figure 4.3.5.2: LK16-16, showing a normal male karyogram.

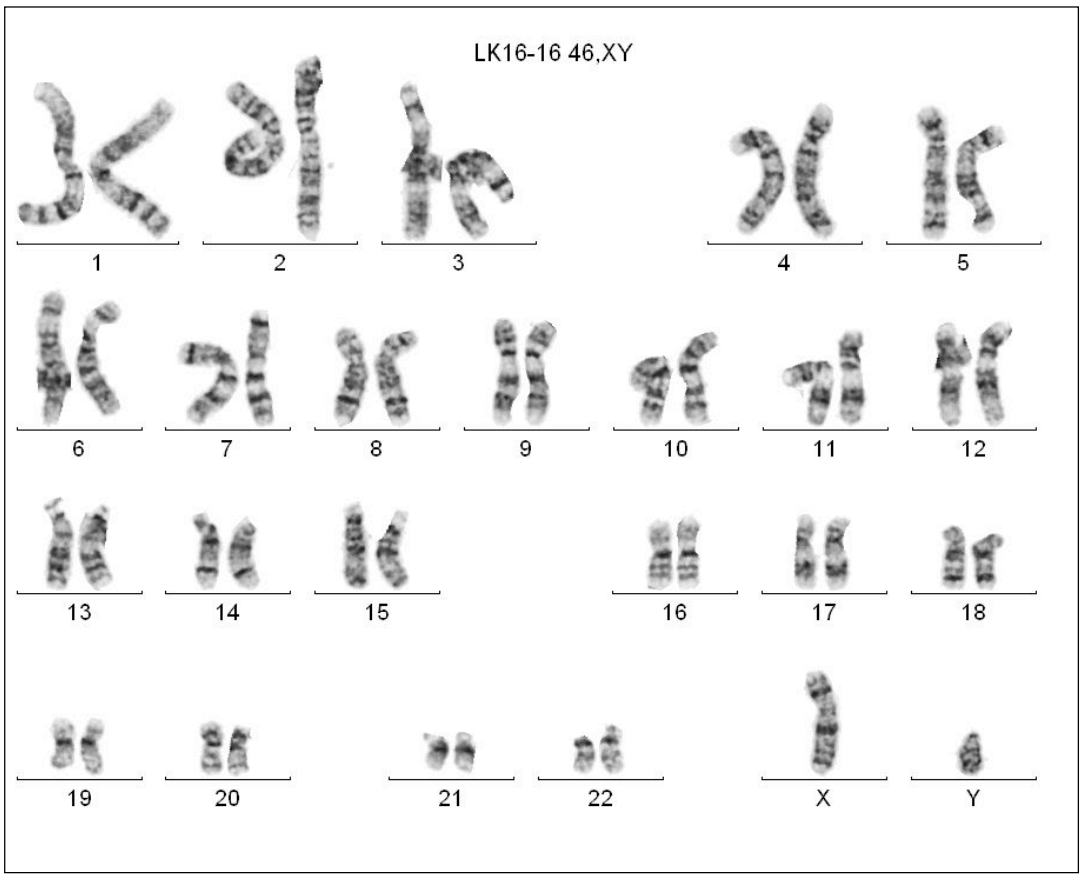


Figure 4.3.5.3: LK6-15, showing a normal male karyogram.

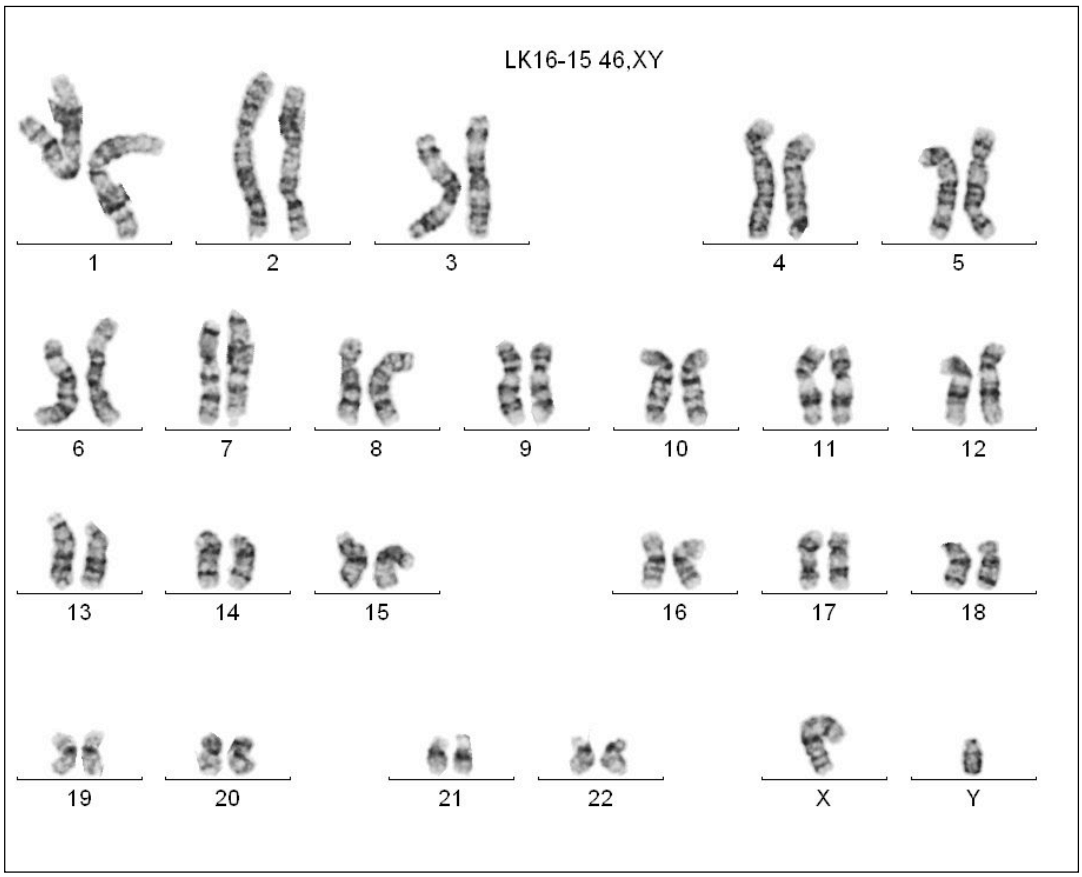


Figure 4.3.5.4 : LK16-15 abnormal cell 46,XY,t(12;14)(q23;q32),del(13)(q12q25).

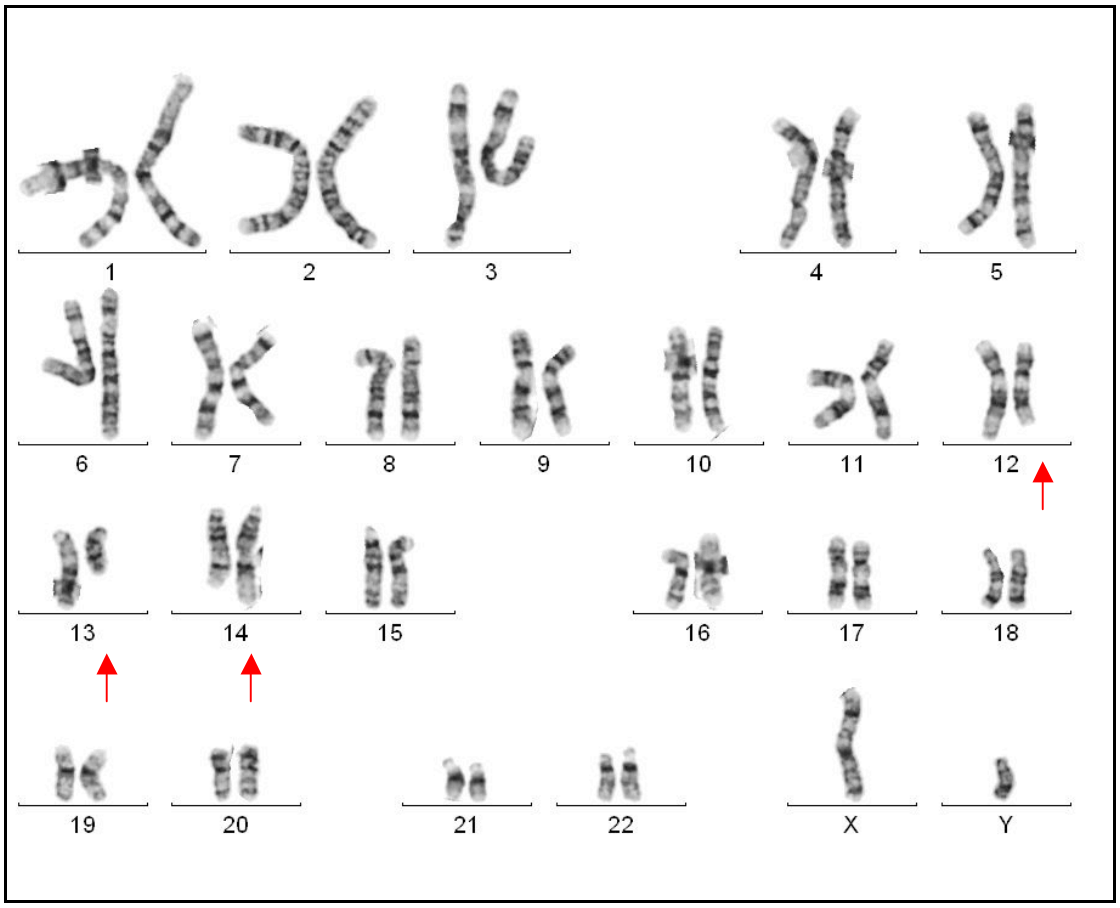
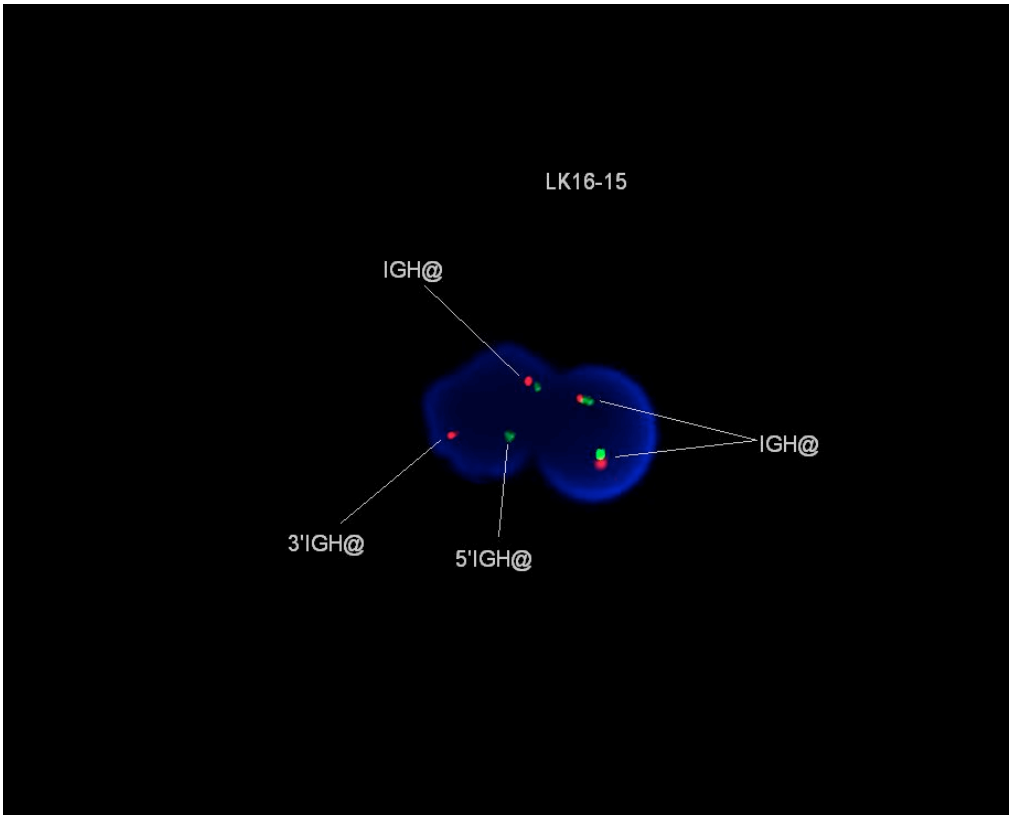


Figure 4.3.5.5: LK16-15 FISH result for IGH@ dual colour break-apart probe.



Two interphase cells are shown. The cell on the right shows two intact IGH@ probes (normal), however the cell on the left shows one intact IGH@ probe and one IGH@ probe that has split into its individual probes (3'IGH@ and 5'IGH@) confirming a rearrangement of this gene.

4.3.6. Laser microdissection of the region of interest on chromosome 15.

To enable next generation sequencing of the entire region of interest on chromosome 15 in this family, laser microdissection using the PALM laser microdissector (Zeiss, Germany) was undertaken on chromosomes from LK16-4 (homozygous for the region of interest), LK16-15 (partial heterozygous) and LK16-16 (full heterozygous). This enabled purification of the region of interest for DNA extraction and amplification using GenomePlex® (Sigma-Aldrich). The amplified DNA was sent to

the Australian Genome Research Facility for next generation sequencing. Figure 4.3.6.1 shows the region that was dissected and Figure 4.3.6.2 shows a before and after photo of the dissection process.

Figure 4.3.6.1: Region dissected 15q14 to 15q22.2

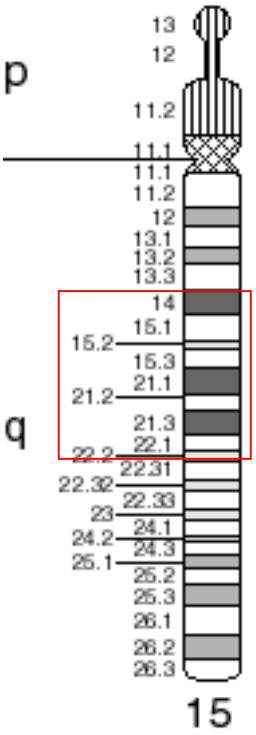
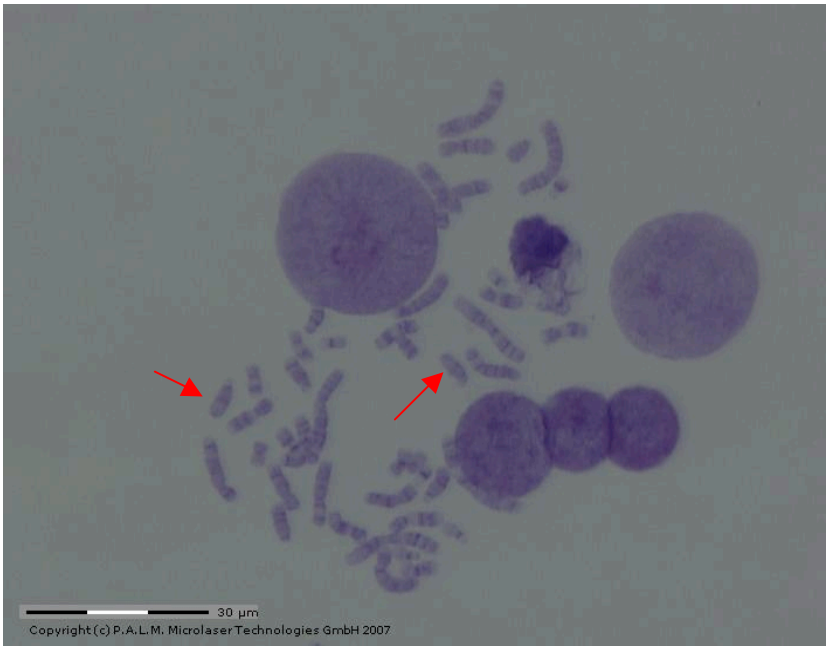
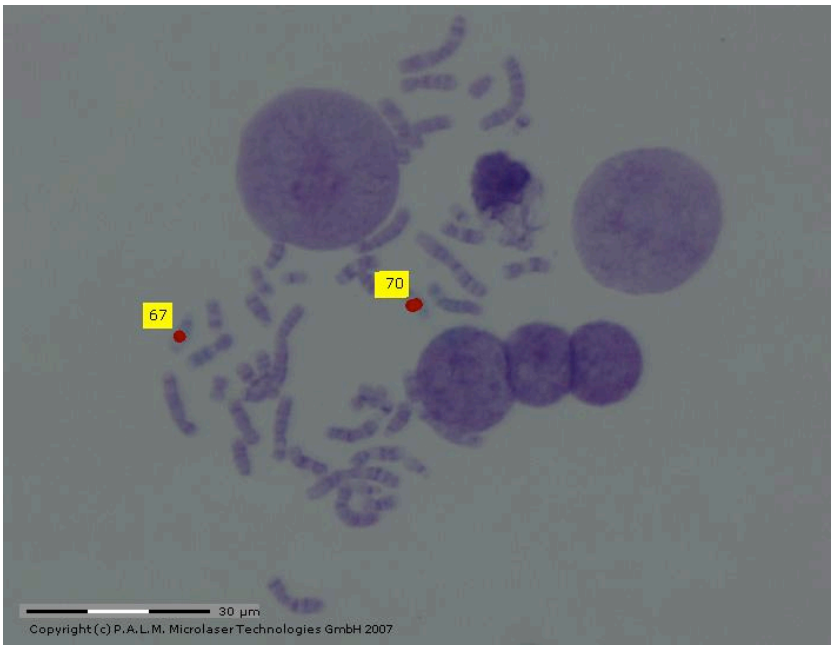


Figure 4.3.6.2: A typical metaphase showing two chromosomes 15

(red arrow) free from cross-overs with other chromosomes prior to dissection (before and after).



The red arrows point to the two chromosomes 15 in this metaphase.



The red dots highlight both chromsomes 15 and show the region under the dot that was dissected off the slide by the laser.

4.4. *Discussion and Conclusions*

The sub-pedigree structure of the LK16 family with five affected siblings from a consanguineous marriage is highly suggestive of a recessively acting mutation with incomplete penetrance. One region in the genome was found where all the affected siblings were homozygous. However it is known that at least one additional mutation has occurred in LK16-1s' genome for the CLL to develop (deletion identified on chromosome 13q14) and that LK16-4 is also a homozygous carrier for the region of interest but is unaffected; he at the age of 83 years has not currently developed a HM. During the course of this study, a reciprocal translocation and a chromosome deletion was discovered in LK16-15, who is the child of LK16-12. The translocation was between the long arm of chromosome 12 and the long arm of chromosome 14 involving the *IGH@* locus, which was confirmed with a fluorescently labelled locus specific identifier break apart probe for the *IGH@*. This translocation has been previously reported to occur in CLL(174). The deletion of the long arm of chromosome 13 is the most common cytogenetic abnormality seen in CLL(175), and this was also present in LK16-1. Further diagnostic testing of LK16-15 confirmed a population of lymphocytes that were co-expressing CD5, CD19, CD20, CD23 and showed an excess of lambda light chains. With the lymphocyte count in this person, this was consistent with MBL. It was intriguing to note that in this person (LK16-15), these cells showed lambda light chain expression as it has been reported in the literature, for MBLs to preferentially express lambda light chains, when there should be more cases of kappa light chain restriction(176) as there are more lymphocytes that express kappa then lambda in a normal person.

MBL is an abnormal monoclonal B lymphocyte population that is considered pre-malignant(1). It is found in approximately 0.3%-3.5% of the general population (176, 177); but in families with CLL the incidence is reported to be increased at 13.5%(177) to 16%(176). Currently, with MBL it is unknown how many of these patients who will subsequently develop CLL(176).

LK16-15 was the person who critically defined the lower limit of the region of interest to 37872929bp. Both LK16-1 and LK16-4 are homozygous on chromosome 15 from 30301633 to 57481797bp. All the other offspring of the siblings (LK16-106, LK16-5, LK16-6, LK16-7, LK16-8, LK16-11 and LK16-16) are heterozygous from 30301633 to 57481797bp.

A genome-wide association study identified 6 risk alleles for CLL(18), and a study by Crowther-Swanepoel *et al*. 2010 identified and additional 4 risk alleles(130). The risk increases for each risk allele a person carries. LK16-1 carries 4 of these risk alleles and therefore has an odds ratio of 2.23. LK16-4 (the unaffected brother of LK16-1, who is homozygous for the chromosome 15 region) carries 5 risk alleles. Interestingly LK16-15, who is the son of LK16-12 and the person whom has MBL and critically defines the lower limit of the region of interest, carries 5 risk alleles; whereas his brother (LK16-16) only carries 3 risk alleles. However it must be noted that that the authors of this paper, Di Bernardo *et al*. 2008(18), estimate that "the six loci we have identified to date through our GWAS account for ~3% of the excess familial risk of CLL". The ten risk alleles identified account for approximately 10% of the heritable risk in CLL(130). Therefore this is a complex disease with many variants involved and it is unlikely that these six loci account for the increased risk in the LK16 family.

DNA extracted from FFPET is prone to degrade and results in fragmentation of the DNA. A recent study found that high-molecular weight DNA can be extracted from FFPET only if it has been stored for less than 6 months(178); and using DNA from FFPET samples older than 6 months greatly decreases the chances of detecting point mutations, deletions and triplet repeat expansions in these samples(178). It has also been reported that assessing microsatellite size in FFPET from tumour samples requires caution as microsatellites can be unstable in tumour tissue and loss of heterozygosity in the tumour tissue may cause false homozygous results(179). The age of the FFPET samples from some of the subjects in my study ranged from several years to decades. The DNA was highly fragmented and was found not to be suitable for SNP array. It was also found that some of the PCR primers for the microsatellite analysis would not hybridise to the fragmented DNA, and the DNA from LK16-12 was the most highly fragmented compared to the other samples. However as the results from the FFPETs could be confirmed by checking for Mendelian errors with siblings and offspring, they could be confidently called in these samples. However DNA from FFPET was crucial to confirm the initial SNP analysis by using DNA extracted from LK16-14, as he was not represented in this initial analysis, as he had no children. The results from the microsatellite analysis of LK16-14 independently helped to confirm the region on chromosome 15 in this family.

From the combined microsatellite and SNP data on both the siblings and offspring it was concluded that the region of interest in this family was chromosome 15q14 (30919517 to 37872929) and that genes in this region should be sequenced. Although it was recognised that LK16-1 and LK16-4 were actually homozygous from

30919517 to 57481797 on chromosome 15. There were 33 genes in the smaller region of homozygosity. There were no obvious candidate gene and no other reports in the literature that implicated this region. Several genes in this region were sequenced, but no mutations found.

It is always possible that one of the siblings developed a sporadic HM, therefore it was decided to dissect the larger region of homozygosity (30919517 to 57481797) on chromosome 15 from chromosomes from LK16-4, LK16-15 and LK16-16, as the region was narrowed to only one nephew (LK16-15). The laser dissection approach was chosen as the method for isolating the target segment of DNA. Next generation sequencing requires enrichment of the target DNA prior to sequencing(180). Currently most studies are doing this by capture-based enrichment methods(155, 181, 182). However, these methods have some limitations: they require the prior knowledge of the sequence that is to be sequenced, they are not suitable for when rare sequence rearrangements are present; they do not include regions that were not sequenced by the Human Genome Project, such as peri centric heterochromatin, regions where there are no known genes (most are exon capture methods); and they are not suitable for when there are mutations corresponding to capture sites (180). Chromosomal dissection of the region of interest avoids all of these problems (180) and was undertaken in this family.

However, the next generation sequencing of the microdissected chromosomes was unsuccessful, and the reads were non-specific. It was thought that this was probably due to the whole genome amplification process, as several microsatellite analysis of DNA from the dissected chromosomes prior to amplification was successful for microsatellites located on the chromosome 15 region and negative for other regions

in the genome. Previously, whole genome amplification has been found to cause artefact nucleotide sequence alterations(183) and fragmentation from the laser dissection.

The study of the LK16 family identified a region of interest where the five affected siblings were homozygous on chromosome 15.

5. Chapter 5: Genome-wide Illumina• 610 Quad Array of the 13 families.

5.1. *Introduction*

Using families with multiple cases of a particular disease is a proven method (linkage analysis) of discovering causative genetic mutations. Mutations in *BRCA1*, *TP53*, *RB1*, were originally found by linkage studies using families enriched for their respective disorders(28, 93, 120). Although there are hundreds of families with HMs in the literature (17), only one mutation in one family has been found (25). However, most of the families with HMs in the literature are relatively small (with only 2-3 affected members)(18).

Many of the families reported in the literature are being revisited (184) and linkage analysis is being carried out with new high-density SNP arrays (18). There has been several genome-wide linkage analysis studies reported in the literature in families with HMs(18, 136-138). The largest of these was by Sellick *et al.* 2007, who studied 206 families with CLL(18). This study was performed in two parts: phase 1 consisted of a genome-wide linkage scan using the GeneChip® Mapping 10Kv131 Xba array containing 11555 SNP markers (Affymetrix); and phase 2 consisted of samples genotyped with over 10204 SNP markers on version 2.0 of the Affymetrix 10K array(18). There were 105 families in phase 1 and 101 families in phase 2, comprising a total of 155 families with CLL and 51 families with CLL and other B lymphoproliferative disorders(18). In total there were 550 people with CLL, NHL or HL(18). This equates to a mean of 2.66 affected subjects per family (550/206). This is in contrast to the 13 Tasmanian families in which there are 130 subjects affected; which equate to a mean of 10 affected per family. Sellick *et al.* 2007, found evidence

of linkage to 2q21.2, 6p22.1 and 18q21.1(18). Fuller *et al.* 2007, found evidence in one family with CLL of linkage to 2q37, 4q35, 11p15 and 14q24(136); whereas Ng *et al.* 2007, found evidence of linkage in 6 families to 13q21.33-22.2(137).

There have been several genome-wide association studies (GWAS) in HMs that have been reported in the literature(144, 145, 147-149, 185, 186); three of these were in CLL(144, 147, 149, 185), two were in childhood ALL(145, 148) and one was in NHL(186). These genome-wide association and linkage studies have identified many genetic regions of interest in HMs, however these loci account for a low percentage of the heritable risk of HMs(130).

The structure of the human genome is more complicated than was first predicted(187-191). It is now realised that immense variation exists and that this ranges from SNPs to large chromosomal abnormalities(191). There are now studies discovering copy number variation (CNV) that range in size from kilobases to megabases(191). These studies have shown that CNVs are heterogenous throughout the genome and that no part of the genome is exempt(191). It has also been found that CNVs can predispose to disease(187); and that most new genome-wide SNP arrays include markers that can be used for CNV analysis.

There are many types of SNP arrays available for use; oligonucleotide arrays (Agilent), Affymetrix SNP arrays and Illumina• SNP arrays (172). The Illumina• 610 Quad array was chosen for the current study as it can be used for both SNP and CNV analysis and it is cost efficient.

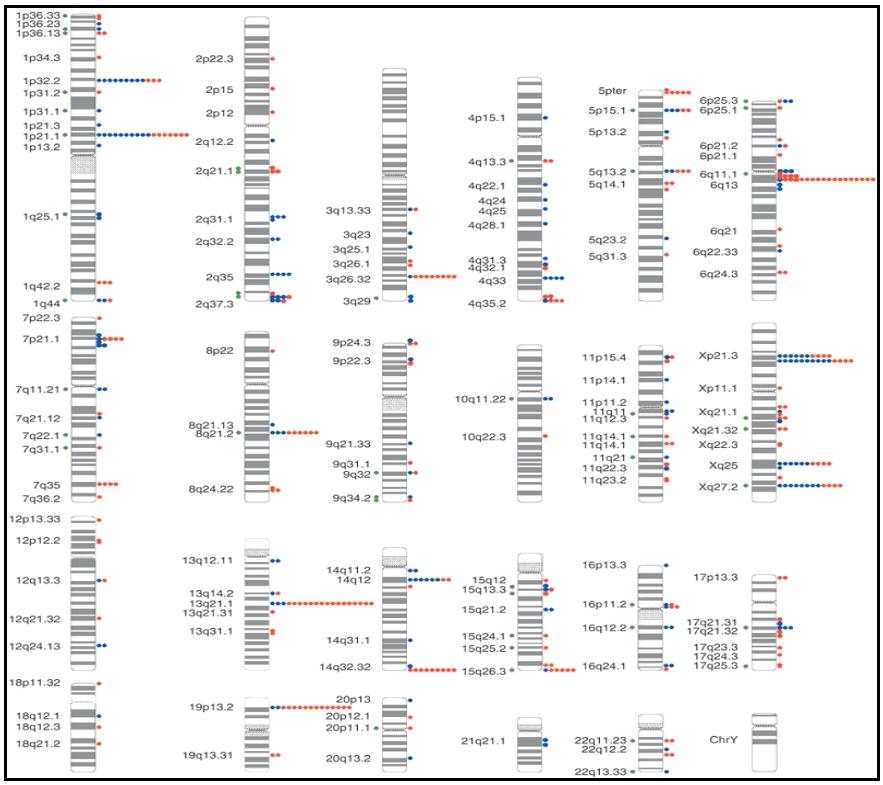
Kidd *et al*. 2008, identified 255 loci across the genome that are sites of CNVs in 39 unrelated individuals(187). They found that CNVs are common to unrelated subjects; and that the CNVs can be in regions where genes are present; and that many CNVs

coincide with segmental duplications in the human genome(187). The CNVs in this paper are shown in Figure 5.1.1.

Genome-wide SNP arrays have also been used on tumour samples to assess CNV and for loss of heterozygosity (82, 154, 172, 192). One study using both a 10K and 50K SNP array on CLL samples found acquired chromosomal abnormalities in 74% of patients, a higher abnormality rate was present on the 50K SNP array compared to the 10K SNP array (192). The most common abnormality found was a deletion of chromosome 13q14 in 53% of patients(192).

The aim of the SNP array in the 13 Tasmanian families was to identify regions of the genome that the affected individuals share in families and across families. This will allow for further investigations of genes that are located in these regions.

Figure 5.1.1: Reported regions of copy number variation



Gains in blue, deletions in red, green circles to the left indicate known genome sequence gaps or known segmental duplications)(187).

5.2. *Materials and Methods*

5.2.1. DNA from alive affected and first-degree relatives of deceased affected family members

Living affected members of the 13 families (as described in Chapter 3) were contacted and asked to provide a blood or saliva sample for DNA extraction (as described in Chapter 2.1). First-degree relatives of deceased affected members were also contacted and asked to participate and to provide blood or saliva for DNA extraction. DNA was extracted as described in Chapter 2.2.

5.2.2. Statistical analysis

Several types of statistical analysis were undertaken on the genotyping data generated using the Illumina• 610 Quad array. Due to the large size of these families and the number of consanguineous marriages, alternative approaches to traditional linkage analysis had to be implemented. These were an identity-by-descent (IBD) sharing approach to identify regions of the genome shared by pairs of cases; a novel family based association analysis; and a standard copy number variation analysis.

IBD Analysis:

This was inferred by Dr Russell Thomson. Identity-by-descent (IBD) between all pairs of cases across all families was inferred. PLINK(193) stripped down the set of SNPs to 51,043 SNPs that are not in linkage disequilibrium (LD) with each other. Allele frequencies and LD were estimated using a sample of 373 Tasmanian unrelated controls. Therefore the number of pairs of cases that shared an allele IBD across the genome (similar to the S_{pairs} statistic(194)) could be generated for each SNP. These statistics followed a Poisson distribution. This property was then used to convert the statistics into p-values.

Family based SNP association:

This association was done by Dr Jac Charlesworth. This tested for significant differences between the allele frequencies of cases and controls, taking into account the relationships between cases. Genotypes for deceased cases were imputed using Merlin-based imputation procedure implemented in PEDSYS(195). Genotypes were not directly imputed; rather genotype covariates were calculated, ranging between 0 and 2, representing the observed or expected number of minor alleles. These numbers were translated back into actual genotypes for integral values of the

covariate (±0.1). These inferred genotypes for the cases were combined with the genotypes of a dataset comprising unaffected Tasmanian controls (n=373). Then tested for association at each SNP by calculating three log likelihoods: firstly the log likelihood of the observed cases genotypes alone; secondly of the observed controls genotypes alone; and thirdly of the combined genotypes. The difference between the combined log likelihood and the sum of the other two log likelihoods is asymptotically distributed as a chi-square variate with 1 degree of freedom.

Copy Number Variation analysis:

This was analysed by Dr Patrick Danoy with PennCNV software. PennCNV implements a hidden Markov model that integrates multiple sources of information (considers SNP allelic ratio distribution as well as signal intensity) to infer CNV calls for individual genotypes samples(196).

5.2.3. Fluorescence *in situ* hybridisation (FISH)

Fosmid probes were ordered from The Hospital for Sick Children in Canada to cover the regions of the genome identified by the CNV analysis in order confirm the number of disease associated alleles in the family members. FISH was performed as described in Chapter 2.10.

5.3. *Results*

5.3.1. Patient recruitment

A total of 258 people from these 13 families consented to participate in this study. An Illumina• 610 Quad SNP array was carried out on 117 subjects from the 13 Tasmanian families. Of these 117 people, 18 previously were affected with a HM, 3 currently had circulating disease cells from a HM, 89 were offspring of a deceased

affected, 2 were a related spouse of an affected and 5 were siblings or parents of a deceased affected. In total, 78 out of 130 affected cases were represented in the genome-wide SNP array (Table 5.3.1). The HMs represented by these 78 is shown in Table 5.3.2.

Table 5.3.1: Subjects submitted for the 610 Illumina• SNP array

	Affected	Offspring of	Disease
LK1-1	Yes		MALT
LK1-5		LK1-3	HLMC
LK1-62		LK1-58	RAEB
LK1-64		LK1-58	
LK1-91		LK1-80	MM
LK1-94		LK1-80	
LK1-103		LK1-100	RAEBII
LK2-11		LK2-9	NHL
LK2-14		LK2-9	
LK2-22		LK2-21	MDS
LK2-24		LK2-21	
LK2-35		LK2-33 (spouse and relative)	CMML
LK2-95		LK2-33	
LK2-97		LK2-1	NHL
LK2-99	Yes		NHL
LK2-131		LK2-30	HCL
LK4-7		LK4-21	DLBCL

LK4-8		LK4-21	
LK4-12	Yes		HL
LK4-37		LK4-39	HL
LK16-1	Yes		CLL
LK16-4		LK16-18	CML
LK16-5		LK16-2	CLL
LK16-11		LK16-9	DLBCL
LK16-15		LK16-12	CLL
LK16-16		LK16-12	
LK16-30		LK16-28	ALL
LK16-31		LK16-28	
LK16-55		LK16-50	DLBCL
LK16-59		LK16-50	
LK16-80		LK16-77	AML M1
LK16-91		LK16-87	CLL
LK16-92		LK16-87	
LK16-104		LK16-101	DLBCL/AML
LK16-105		LK16-101	
LK16-108	Yes		MM
LK26-3		LK26-1	PV
LK26-9		LK26-8	PV
LK26-10		LK26-8	
LK26-19		LK26-18	DLBCL

LK26-20		LK26-18	
LK26-23		LK26-6	FL
LK26-24		LK26-6	
LK51-1	Yes		TNHL
LK51-66		LK51-63	AML
LK51-67		LK51-63	
LK51-73		LK51-70	HL
LK51-74		LK51-70	
LK51-75		LK51-70	
LK51-79		LK51-78	AML
LK51-93		LK51-90	HL
LK51-94		LK51-90	
LK51-99		LK51-97	NHL
LK51-110		LK51-106	NHL
LK51-139		LK51-135	CML
LK51-140		LK51-135	
LK51-158		LK51-065	DLBCL
LK51-159	Yes		DLBCL
LK65-28		LK65-22	DLBCL
LK65-33		LK65-21	MM
LK65-34		LK65-21	
LK124-15		Spouse LK124-14 (3^{rd} cousins)	AML
LK124-16		LK124-14	

LK124-17		LK124-14	
LK124-18		LK124-14 (different mother, not 15)	
LK124-19		LK124-14 (different mother, not 15)	
LK124-37		LK124-35	CMML
LK124-43		LK124-35	
LK124-86		LK124-82	DLBCL
LK124-87		LK124-82	
LK124-95		LK124-93	AML
LK124-117	Yes		AML
LK124-140		LK124-145	DLBCL
LK124-141		LK124-145	
LK124-165		LK124-161	DLBCL
LK124-179	Yes		DLBCL
LK124-202	Yes		APML
LK124-209		LK124-187	DLBCL
LK132-16		LK132-14	CLL
LK132-17		LK132-14	
LK132-35		LK132-31	MM
LK132-37		LK132-31	
LK153-1	Yes		DLBCL
LK153-3	Yes		MCL

LK537-1	Yes		CLL
LK537-2	Yes		CLL
LK836-1	Yes		CLL
LK836-2	Yes		HCL
LK836-6		LK836-4	NHL
LK836-8		LK836-5 (mother)	ALL
LK836-30		LK836-4	
LK836-31		LK836-4	
LK836-53		LK836-50 (sister)	ALL
LK836-54		LK836-5 (brother)	
LK2042-3	Yes		PTALL
LK2042-5	Yes		PBALL
LK2042-6	Yes		HLNS
LK2042-18		LK2042-16	PV
LK2042-19		LK2042-16	
LK2042-20		LK2042-16	
LK2042-29		LK2042-11	WM
LK2042-30		LK2042-11	
LK2042-50		LK2042-8 (father of)	PTALL
LK2042-56		LK2042-2	NHL
LK2042-57		LK2042-2	
LK2042-162		LK2042-7	MM
LK2042-163		LK2042-7	

LK2042-129	Yes		MM
LK2042-180		LK2042-134	MM
LK2042-186		LK2042-9	ET
LK2042-187		LK2042-9	
LK2042-197		LK2042-12	DLBCL
LK2042-198		LK2042-12	
LK2042-203		LK2042-200 (mother of)	PV
LK2042-212		LK2042-207	PBALL
LK2042-231	Yes		DLBCL
LK2042-273		LK2042-270	AML

Table 5.3.2: Distribution of the HM subtypes genotyped in each family.

	MPN	MPN/MDS	MDS	AML	ALL	MBCN	CLL	MTCN	HL
LK1			2			2			1
LK2		1	1			4			
LK4						1			2
LK16	1			1	1	4	4		
LK26	2					2			
LK51	1			2		4		1	2
LK65						2			
LK124		1		4		5			
LK132						1	1		
LK153						2			
LK537							2		
LK836					2	2	1		
LK2042	3		1	1	4	6			1
Total	**7**	**2**	**4**	**8**	**7**	**35**	**8**	**1**	**6**

5.3.2. Quality review of the genotyping data.

Missingness

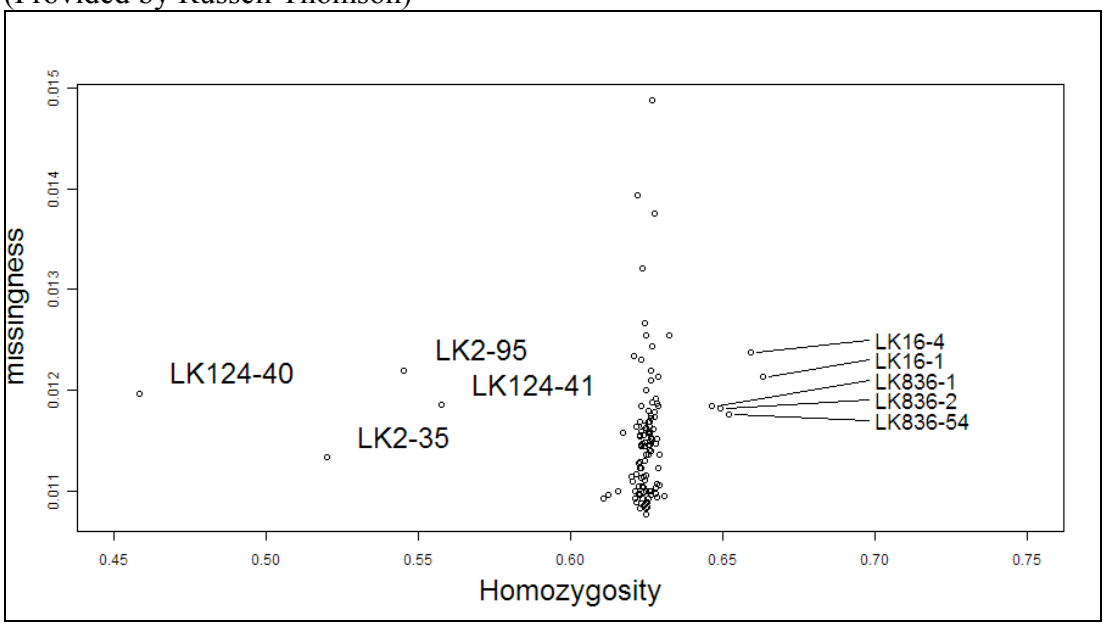
This was assessed with PLINK, and represents the number of SNPs that are not assessable. Ideally a missingness rate should be less than 5%(197). All the samples were under this limit (Figure 5.3.2.1).

Homozygosity

Samples LK16-4, LK16-1, LK836-1, LK836-2 and LK836-54 all had an increased proportion of homozygous calls (Figure 5.3.2.1). There are three explanations for this: firstly that there may be a region of uniparental disomy; secondly that there may be an acquired deletion in the tumour cells; or thirdly that the parents of the individuals are related. All of these subjects were products of consanguineous marriages and represented two families only.

Figure 5.3.2.1: Missingness and homozygosity assessment of the data

(Provided by Russell Thomson)



5.3.3. Copy number variants (CNVs)

Figure 5.3.3.1a,b,c demonstrates the total number of CNVs present in each persons genome. There was a total of 4276 CNVs detected in these 117 subjects, of which 695 were unique. The median number of CNVs detected in each person's genome was 36.53 (range 17-92). Figure 5.3.3.2 shows all the 4276 CNVs and their type and location in the genome. Each row on the Y-axis on this graph represents an individual persons' genome and along the X-axis is each chromosome (1 to 22) represented from p arm to q arm along the axis. A copy number change of 0 (homozygous deletion) is represented by a red line at its location, a CNV of 1 (heterozygous deletion) is represented by a green line at its location, a CNV of 3 (heterozygous duplication) is represented by a dark blue line at t's location, a CNV of 4 (homozygous duplication) is represented by a aqua line at its location and where there is no coloured lines the CNV is two (normal number). The specific CNVs by chromosome position are given in Appendix C for all 117 people. It can be seen that there are certain regions in the genome that are very polymorphic, especially a region on chromosome 4, chromosome 6 and chromosome 11; where virtually every sample in this analysis is polymorphic at these sites. Figure 5.3.3.3 shows the CNVs for all the affected individuals, although not one is shared by all affected.

Figure 5.3.3.1a: Total number of CNVs present in each person in the LK1, LK2, LK4, LK16 and LK26 Families.

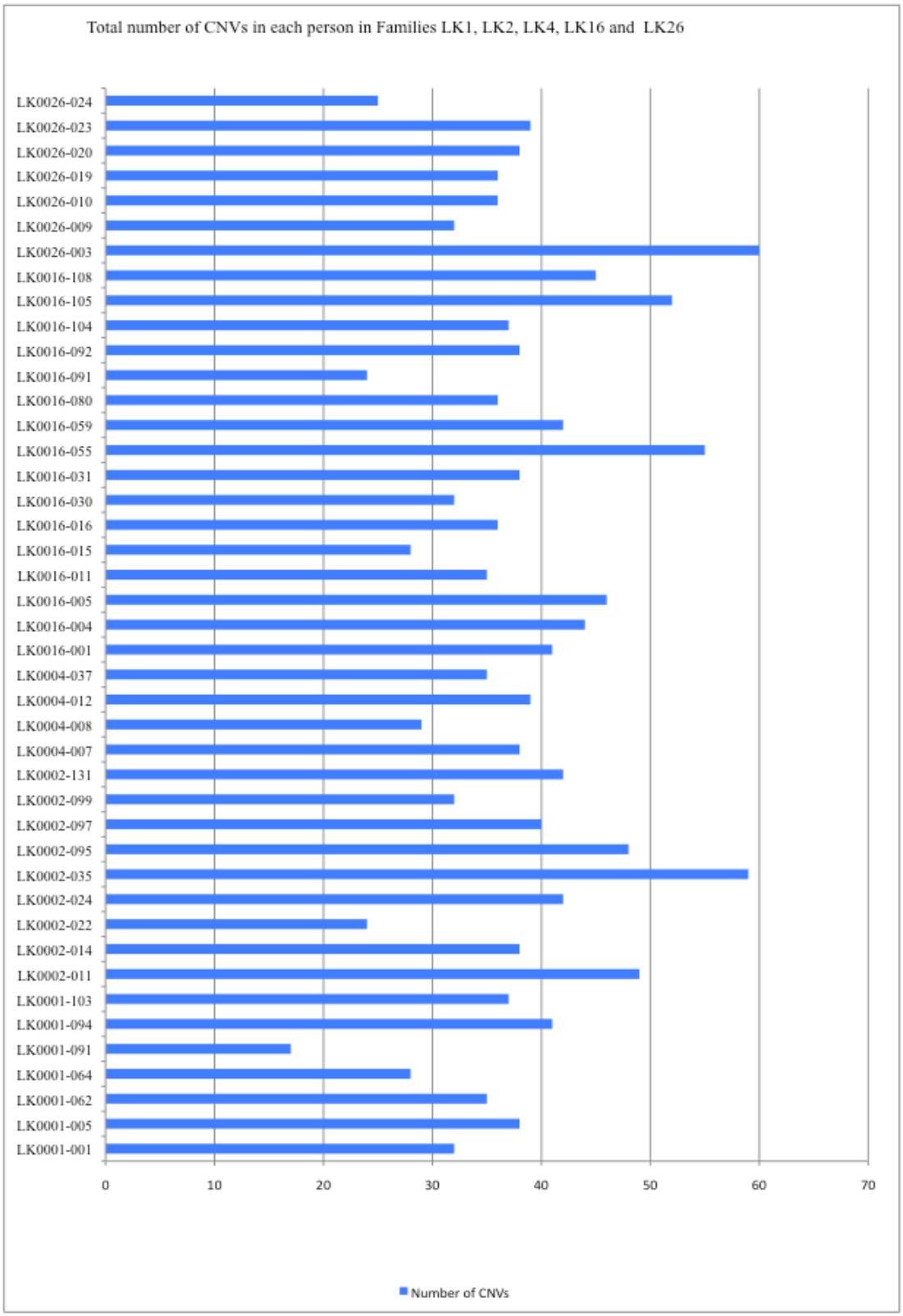


Figure 5.3.3.1b: Total number of CNVs present in each person in the LK51, LK124 and LK132 Families.

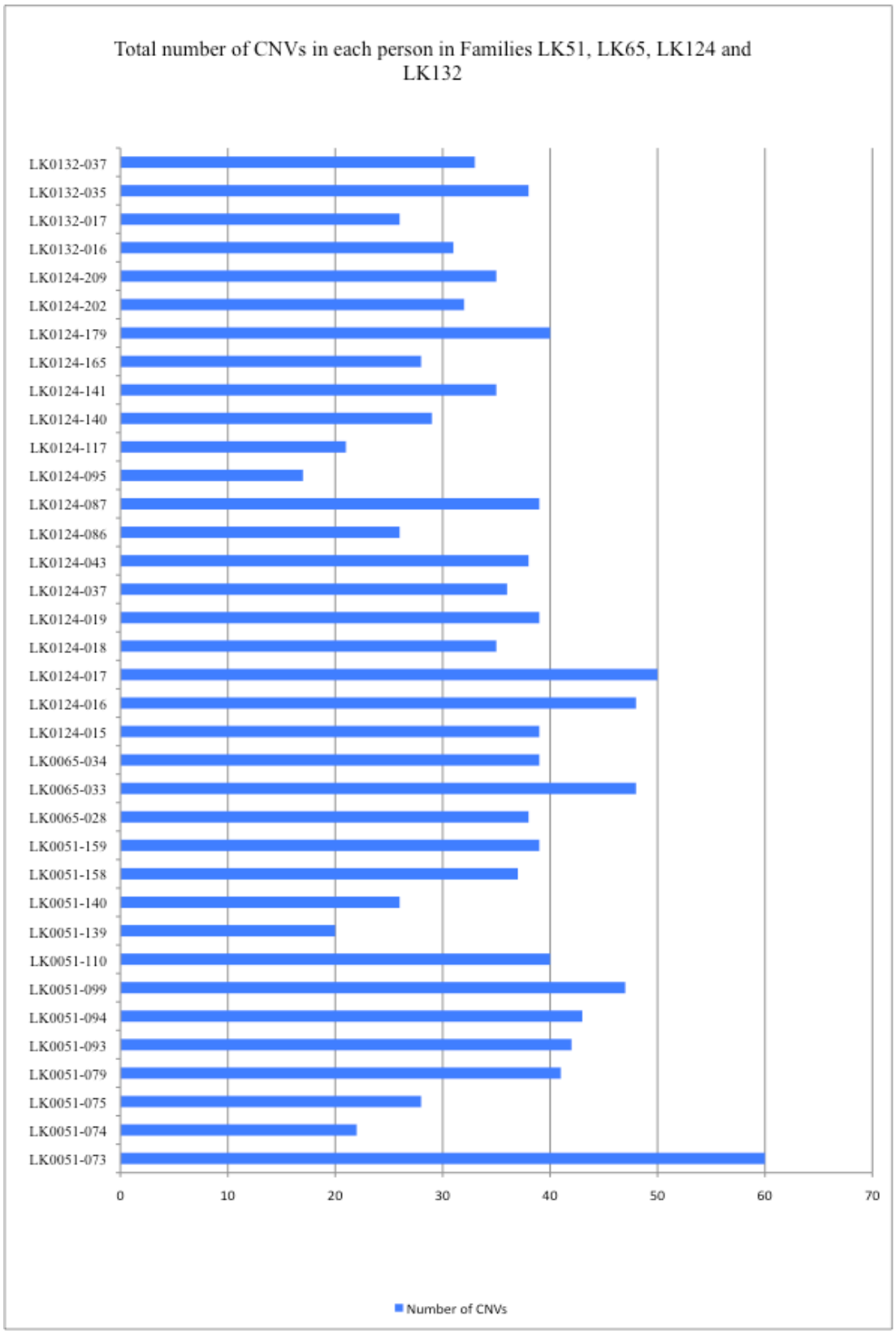


Figure 5.3.3.1c: Total number of CNVs present in each person in the, LK153, LK537, LK836 and LK2042 Families.

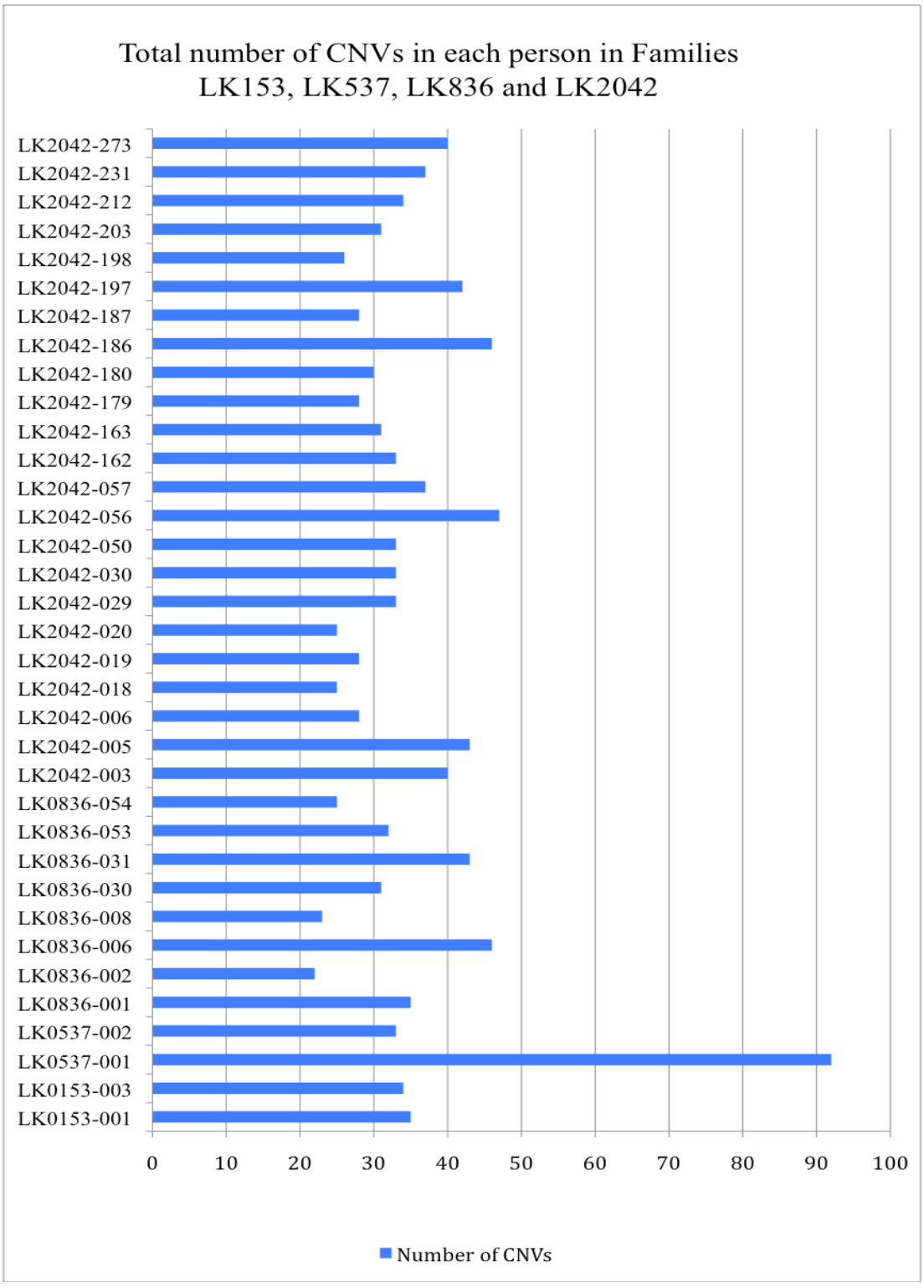


Figure 5.3.3.2: Shows the location and type of CNV for all individuals.

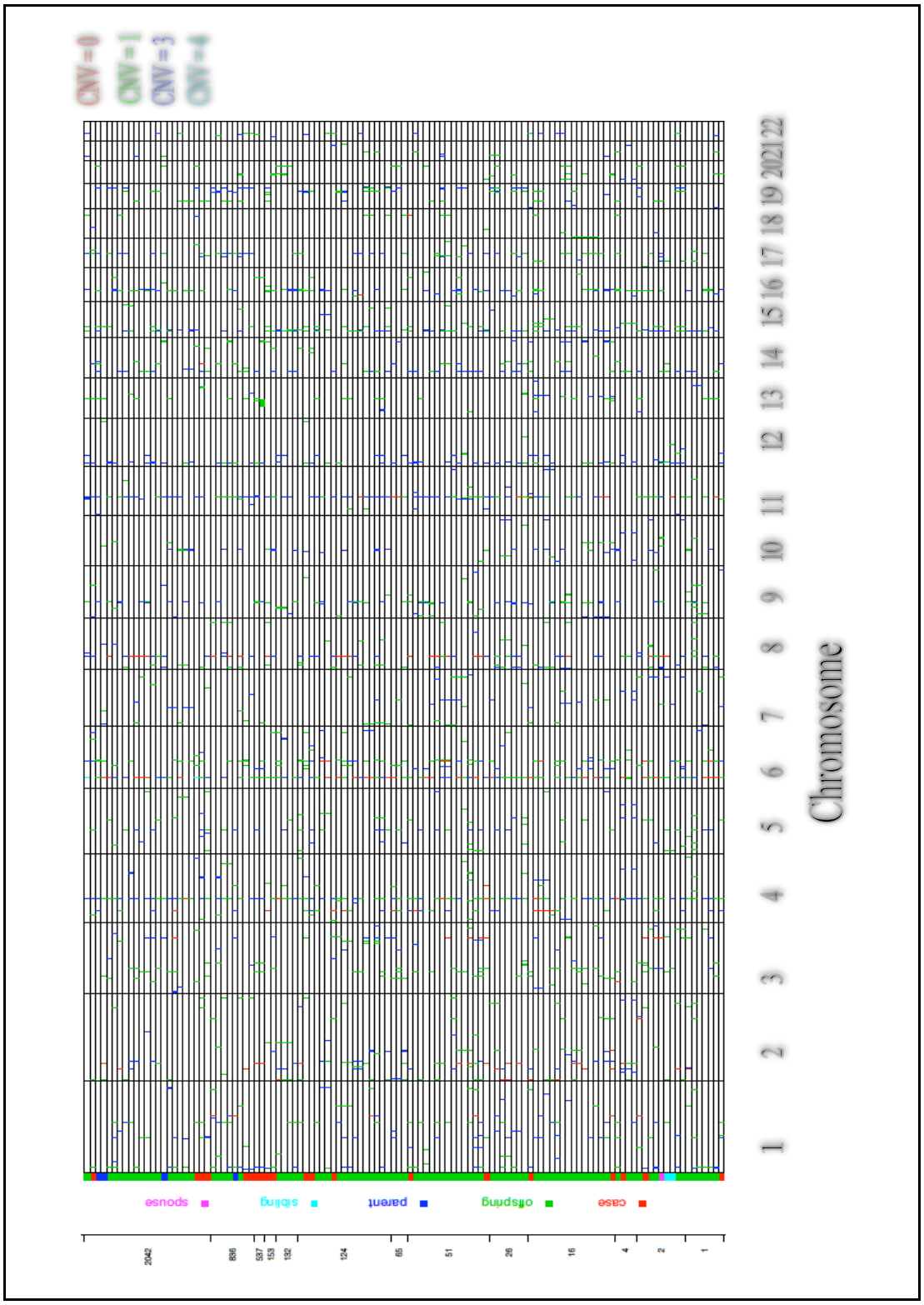
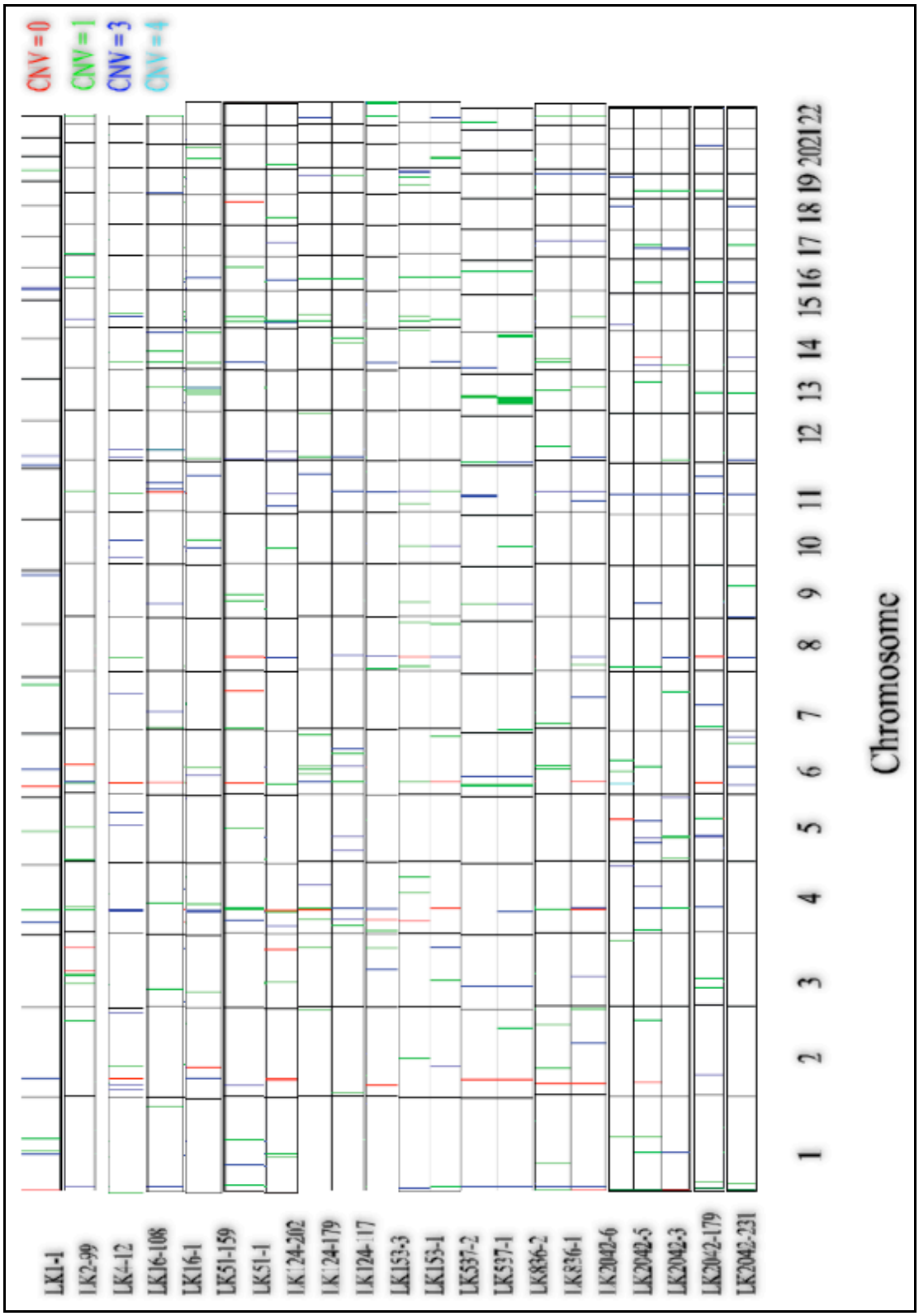


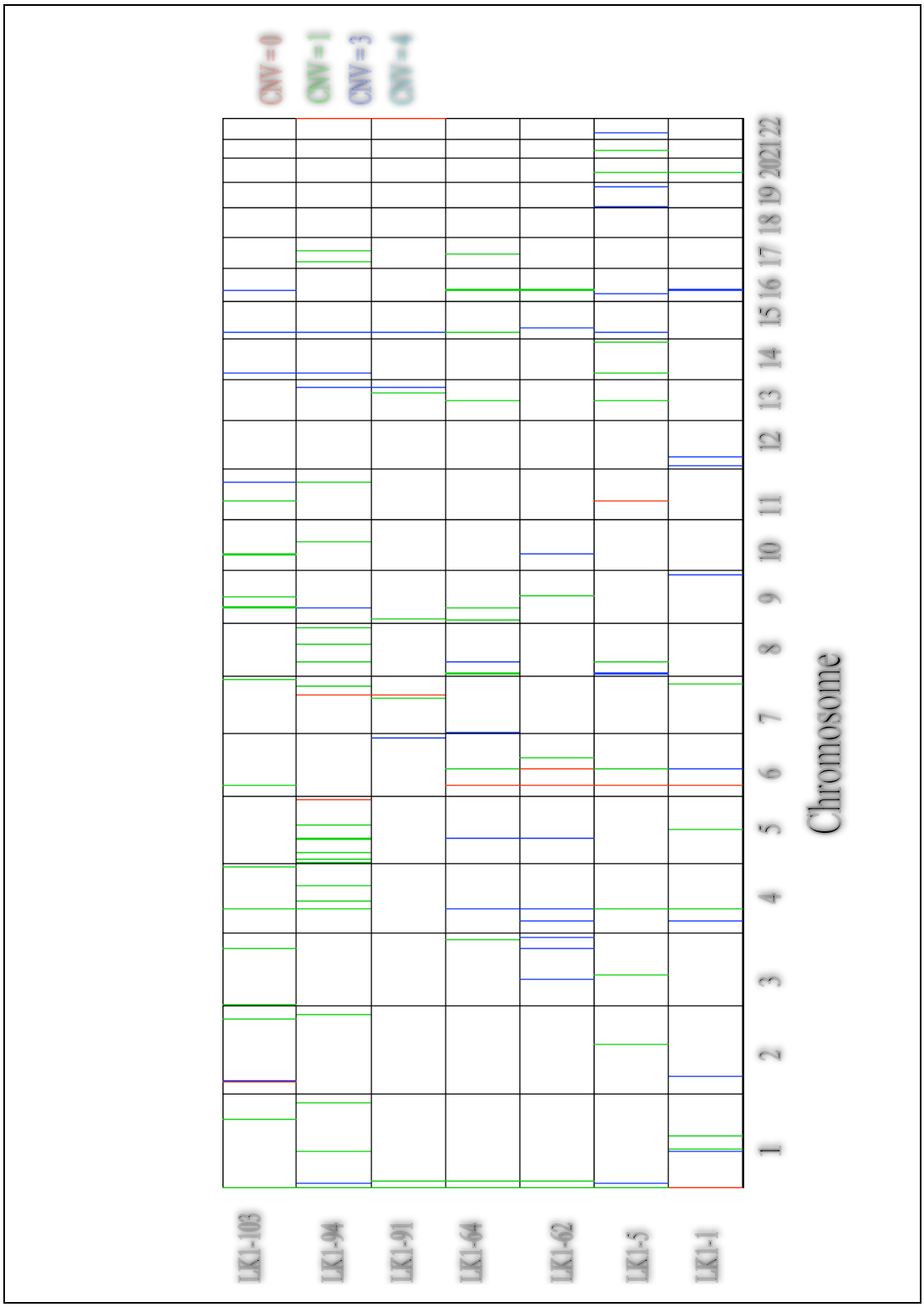
Figure 5.3.3.3: Shows the location and type of CNV for individuals who have been diagnosed with a HM.



5.3.3.1. CNVs in the LK1 family

Seven subjects from this family were included in this study. One person (LK1-1) was affected with a HM, the other six were offspring of another four affected members (see Table 5.3.1). The CNVs in this family are shown in Figure 5.3.3.1.1.

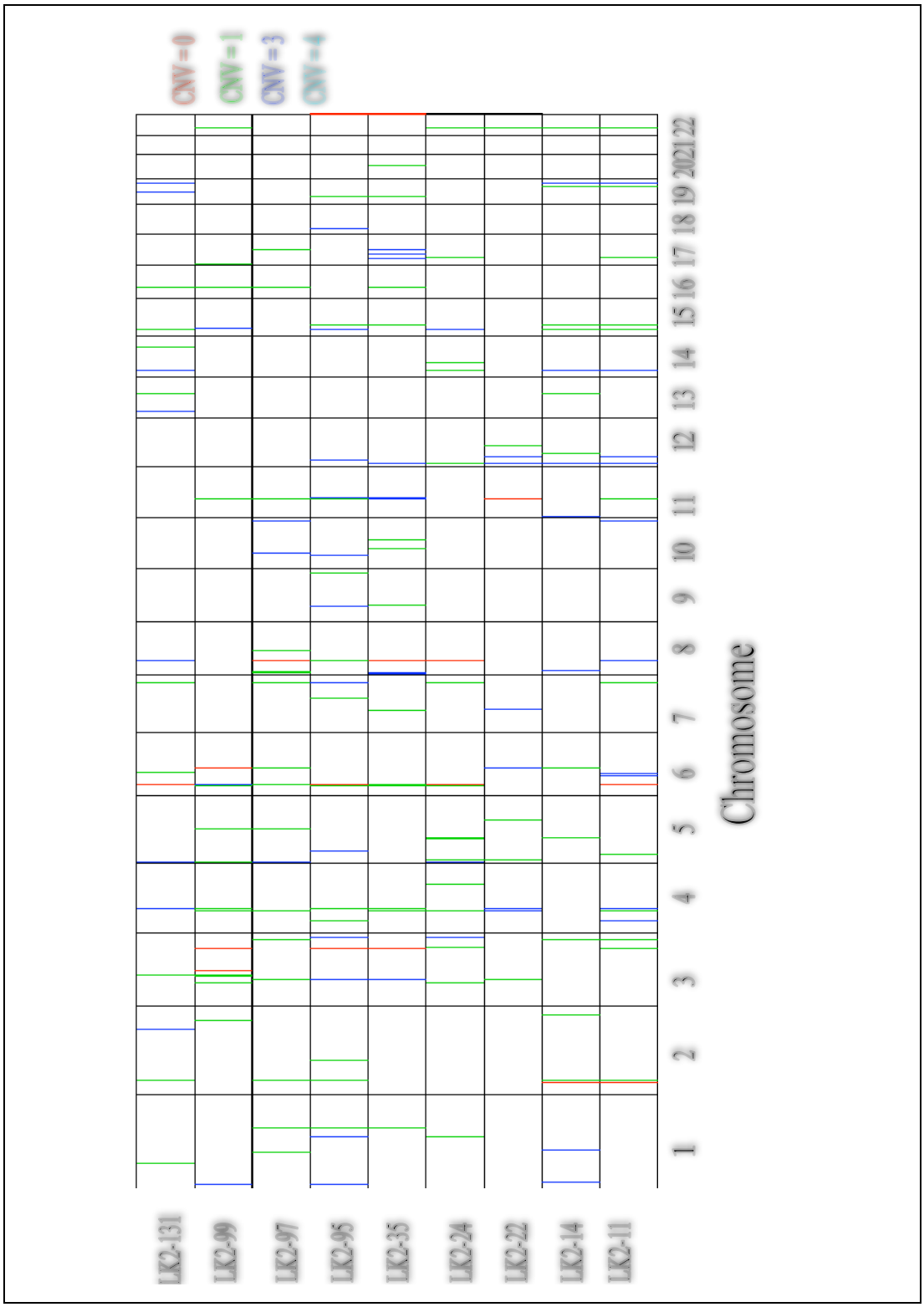
Figure 5.3.3.1.1: Shows the location and type of CNV in the LK1 family.



5.3.3.2. CNVs in the LK2 family

Nine subjects from this family were included in this study. One person (LK2-99) was affected with a HM, the other eight were offspring of another six affected members (see Table 5.3.1). The CNVs in this family are shown in Figure 5.3.3.2.1.

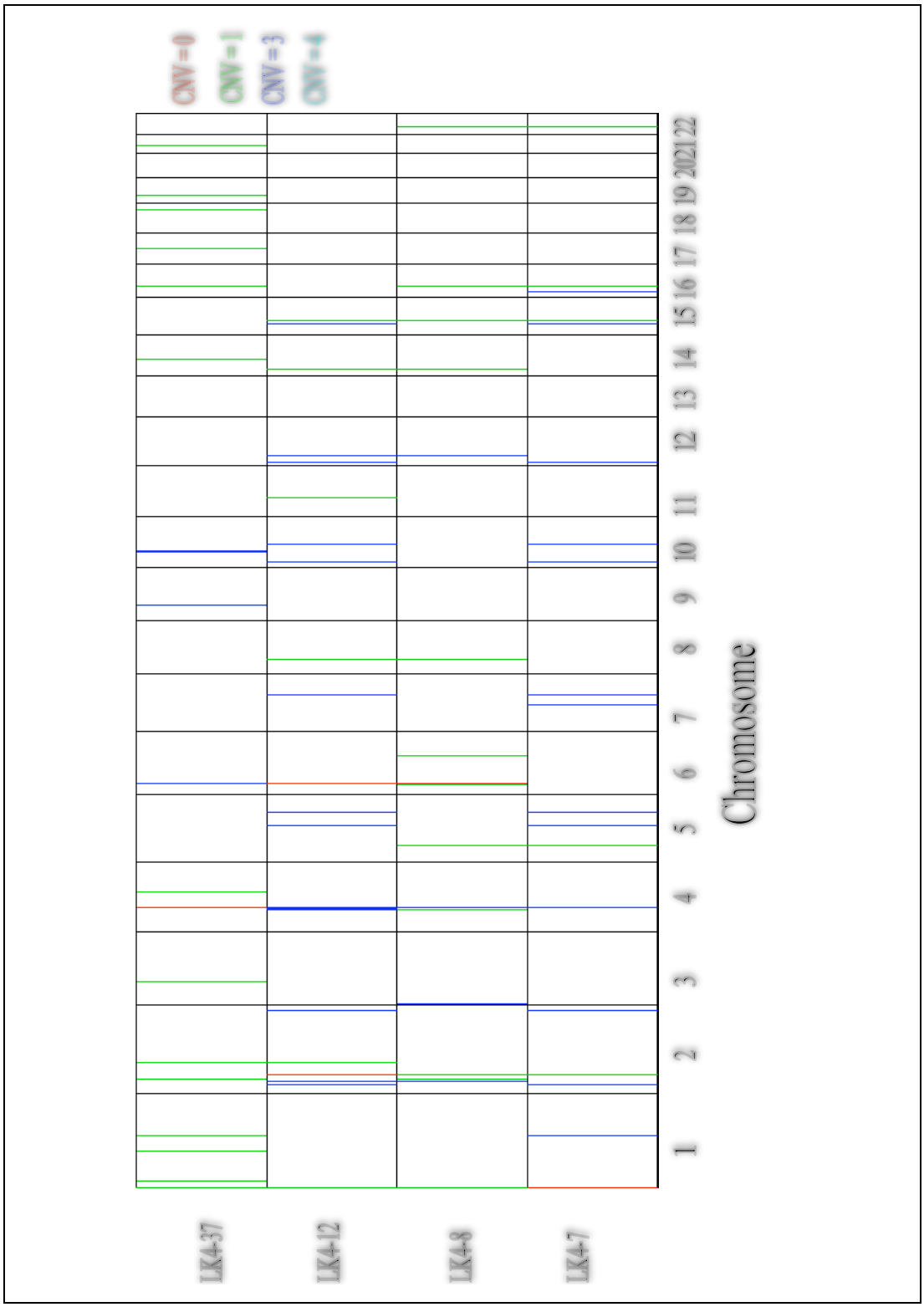
Figure 5.3.3.2.1: Shows the location and type of CNV in the LK2 family.



5.3.3.3. CNVs in the LK4 family

Four subjects from this family were included in this study. One person (LK4-12) was affected with a HM, the other three were offspring of another two affected members (see Table 5.3.1). The CNVs in this family are shown in Figure 5.3.3.3.1.

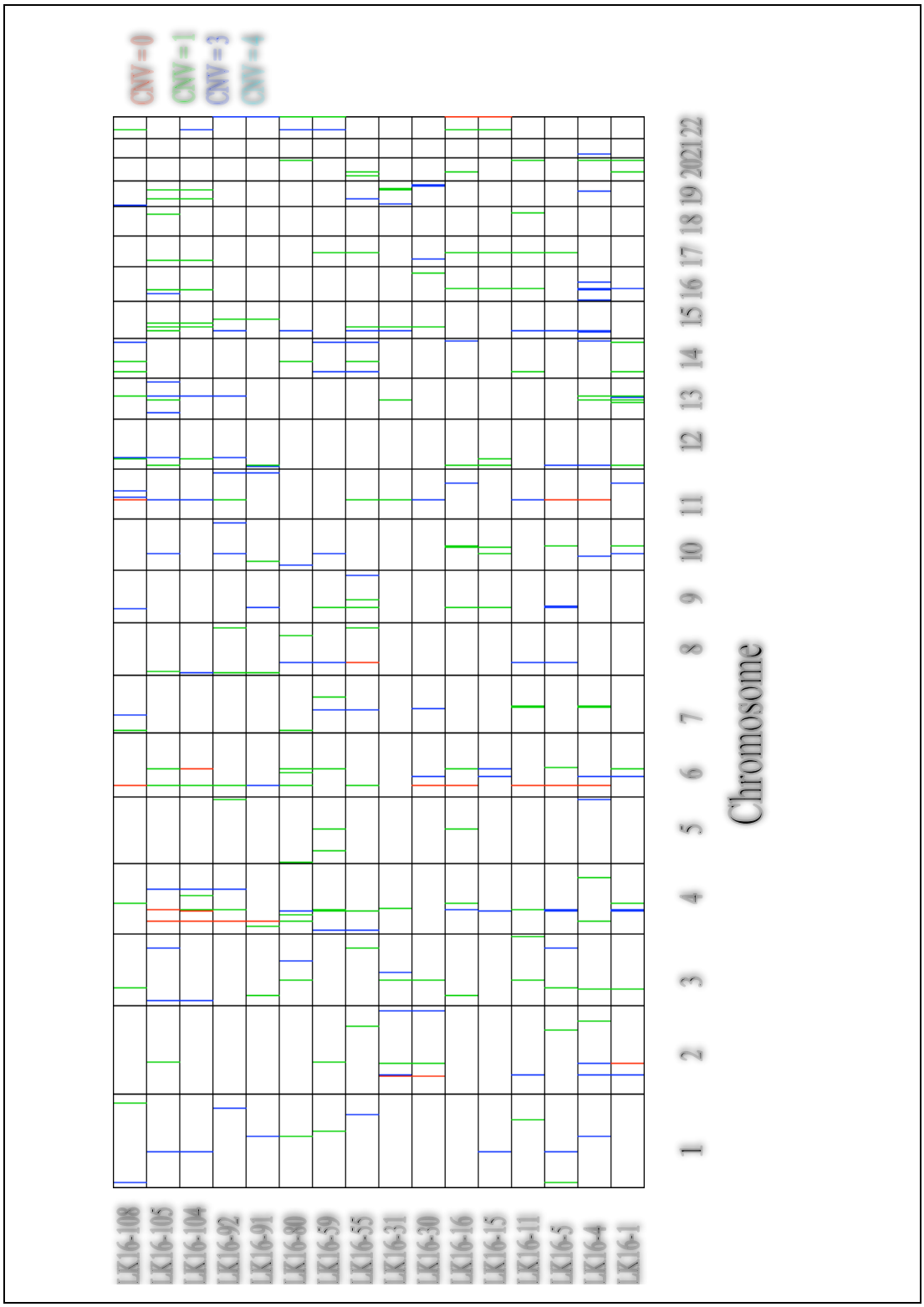
Figure 5.3.3.3.1: Shows the location and type of CNV in the LK4 family.



5.3.3.4. CNVs in the LK16 family

Sixteen subjects from this family were included in this study. Two subjects (LK16-1 and LK16-108) were affected with a HM, the other fourteen were offspring of another nine affected members (see Table 5.3.1). The CNVs in this family are shown in Figure 5.3.3.4.1.

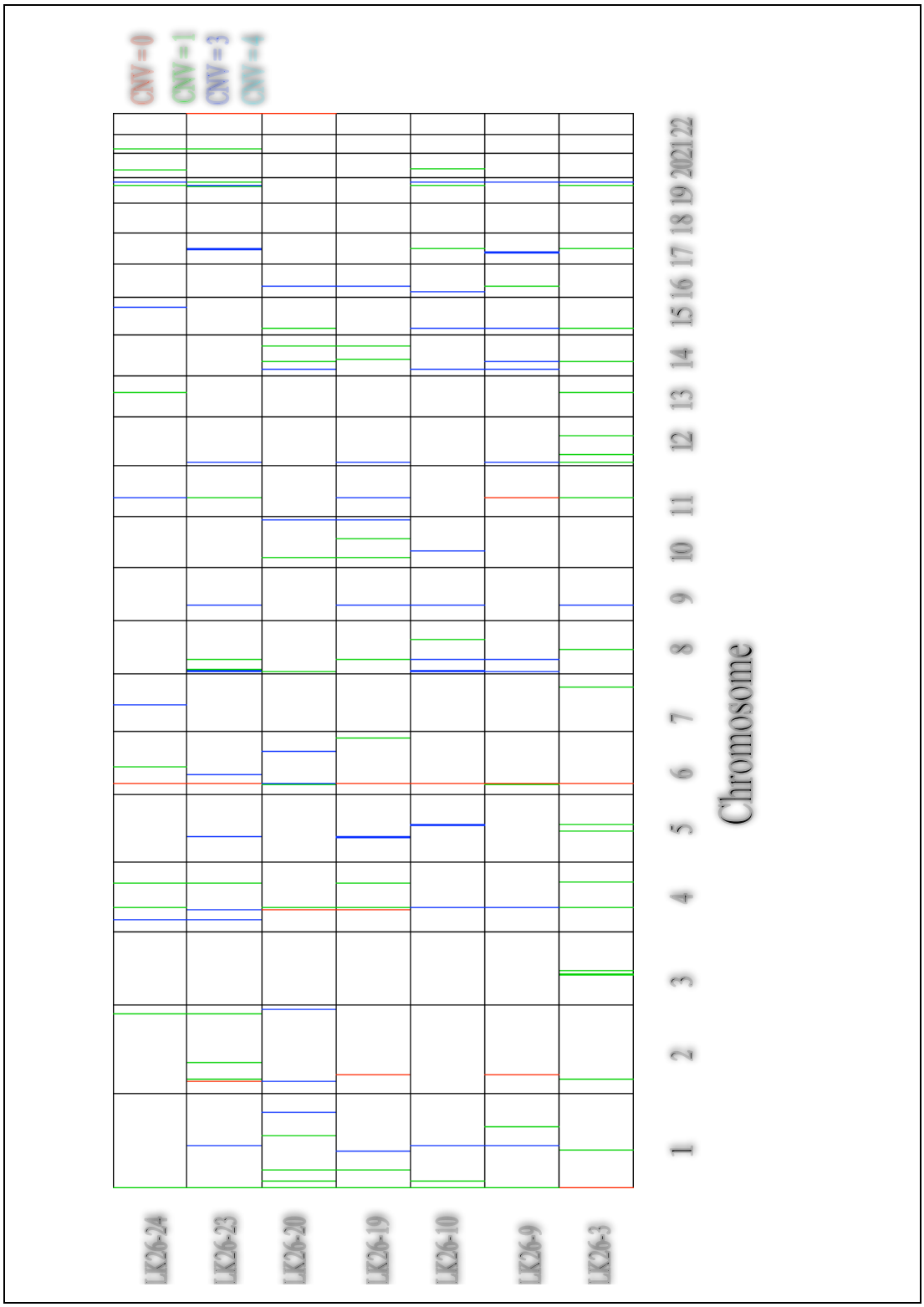
Figure 5.3.3.4.1: Shows the location and type of CNV for the LK16 family.



5.3.3.5. CNVs in the LK26 family

Seven subjects from this family were included in this study. All seven were offspring of four affected members (see Table 5.3.1). The CNVs in this family are shown in Figure 5.3.3.5.1.

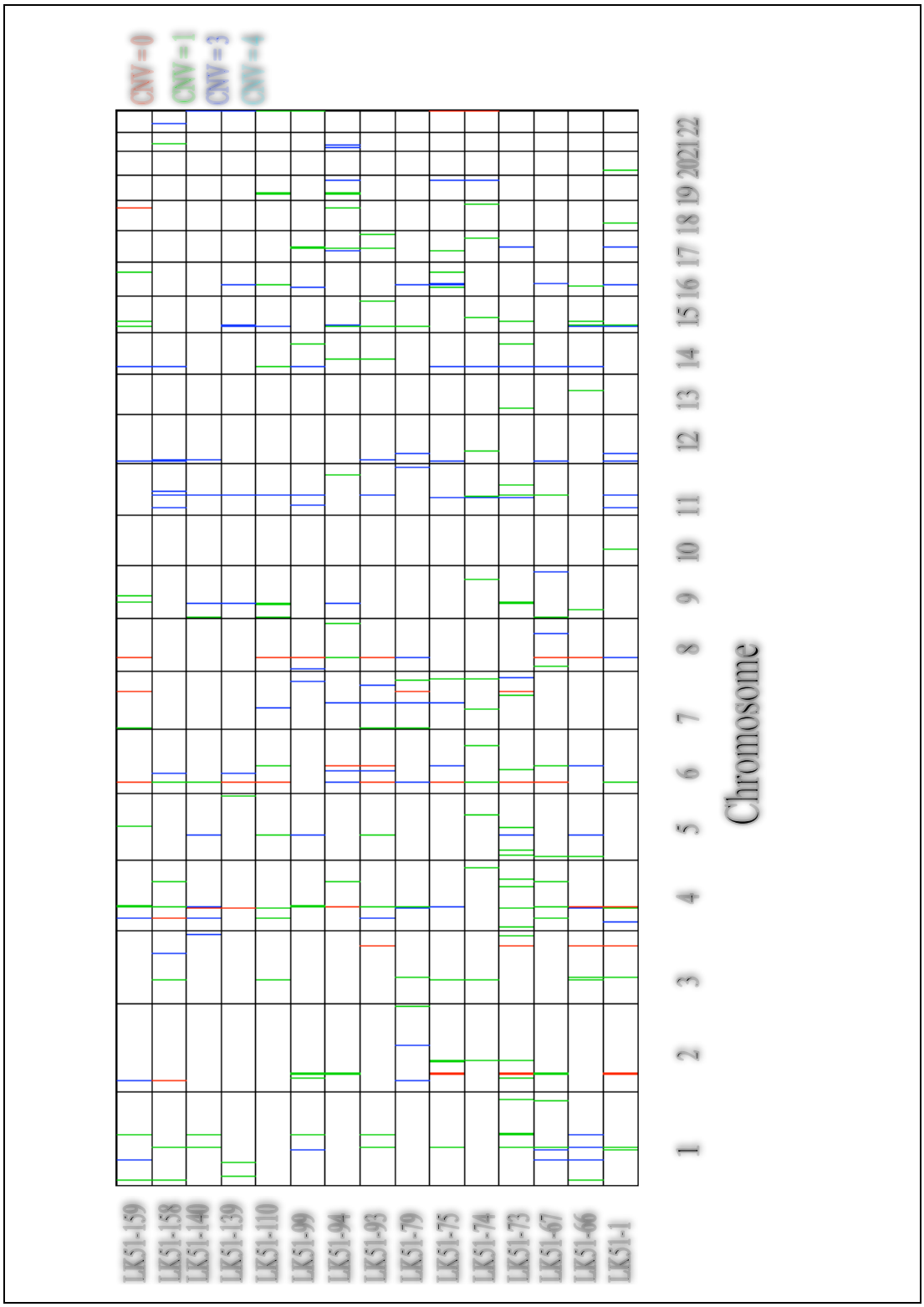
Figure 5.3.3.5.1: Shows the location and type of CNV for the LK26 family.



5.3.3.6. CNVs in the LK51 family

Fifteen subjects from this family were included in this study. Two subjects (LK51-1 and LK51-159) were affected with a HM, the other thirteen were offspring of another eight affected members (see Table 5.3.1). The CNVs in this family are shown in Figure 5.3.3.6.1.

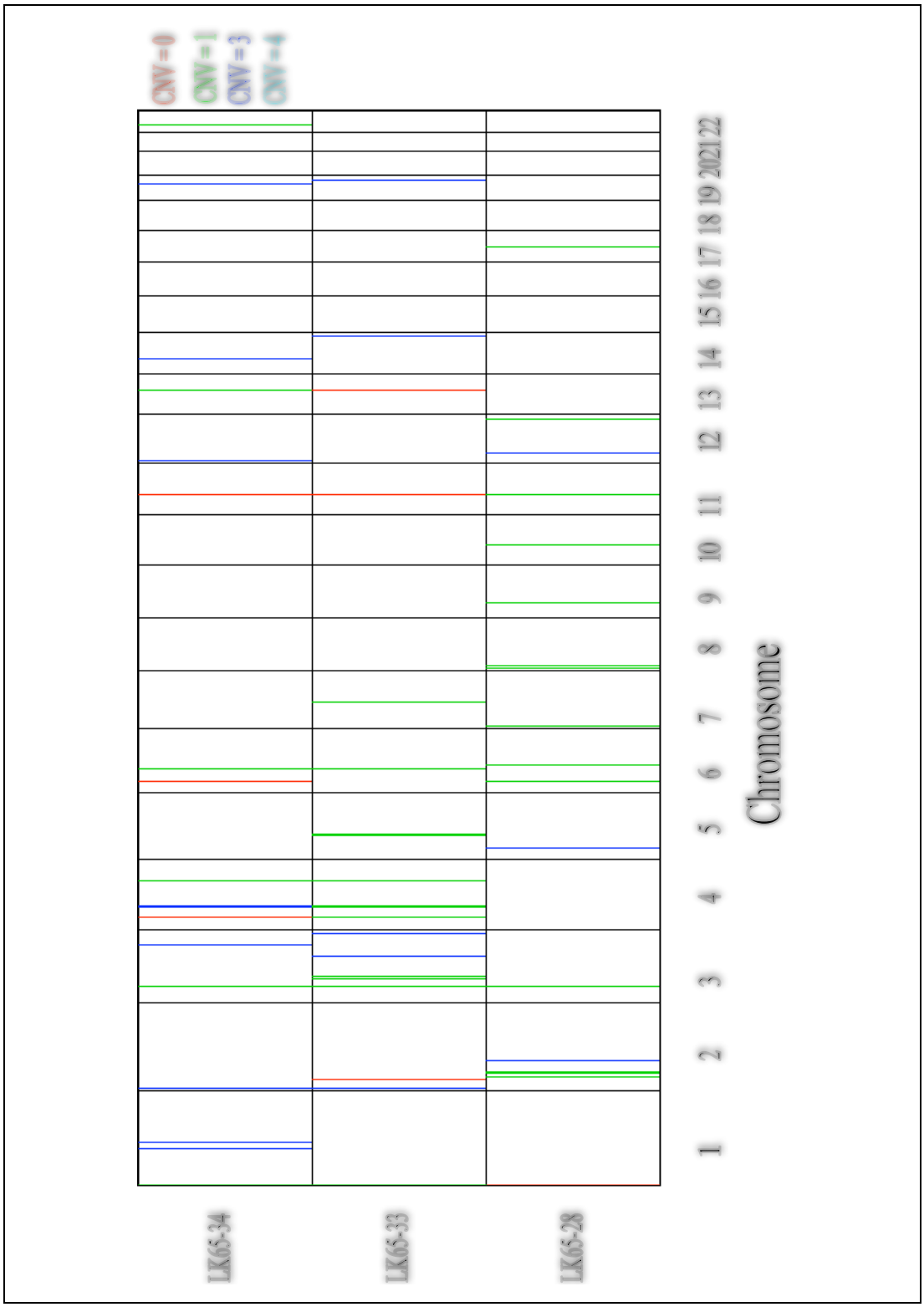
Figure 5.3.3.6.1: Shows the location and type of CNV for the LK51 family.



5.3.3.7. CNVs in the LK65 family

Three subjects from this family were included in this study. All were offspring of two affected members (see Table 5.3.1). The CNVs in this family are shown in Figure 5.3.3.7.1.

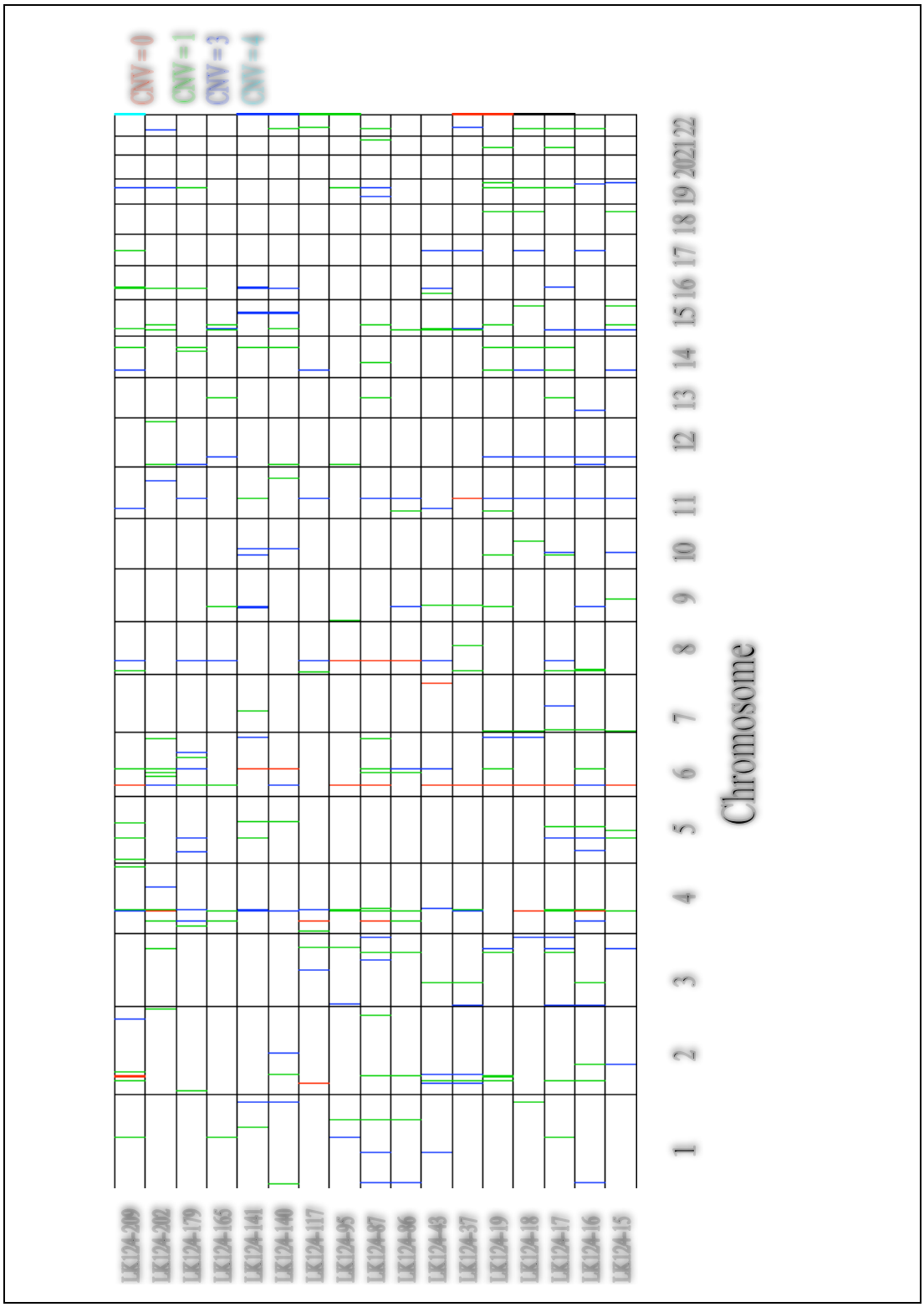
Figure 5.3.3.7.1: Shows the location and type of CNV for the LK65 family.



5.3.3.8. CNVs in the LK124 family

Seventeen subjects from this family were included in this study. Three subjects (LK124-117, LK124-179 and LK124-202) were affected with a HM, the other fourteen were first-degree relatives of another seven affected members (see Table 5.3.1). The CNVs in this family are shown in Figure 5.3.3.8.1.

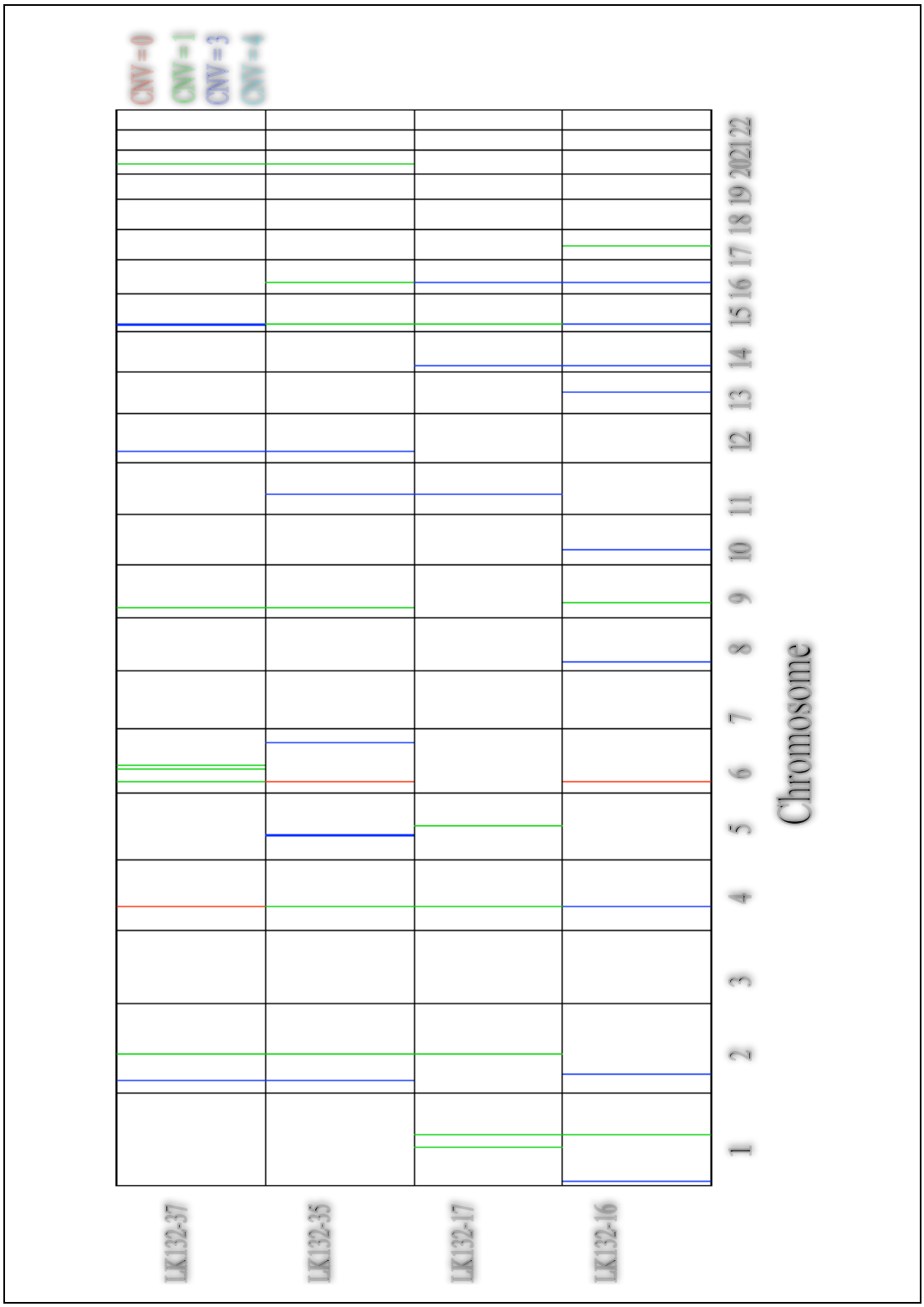
Figure 5.3.3.8.1: Shows the location and type of CNV for the LK124 family.



5.3.3.9. CNVs in the LK132 family

Four subjects from this family were included in this study. All were offspring of two affected members (see Table 5.3.1). The CNVs in this family are shown in Figure 5.3.3.9.1.

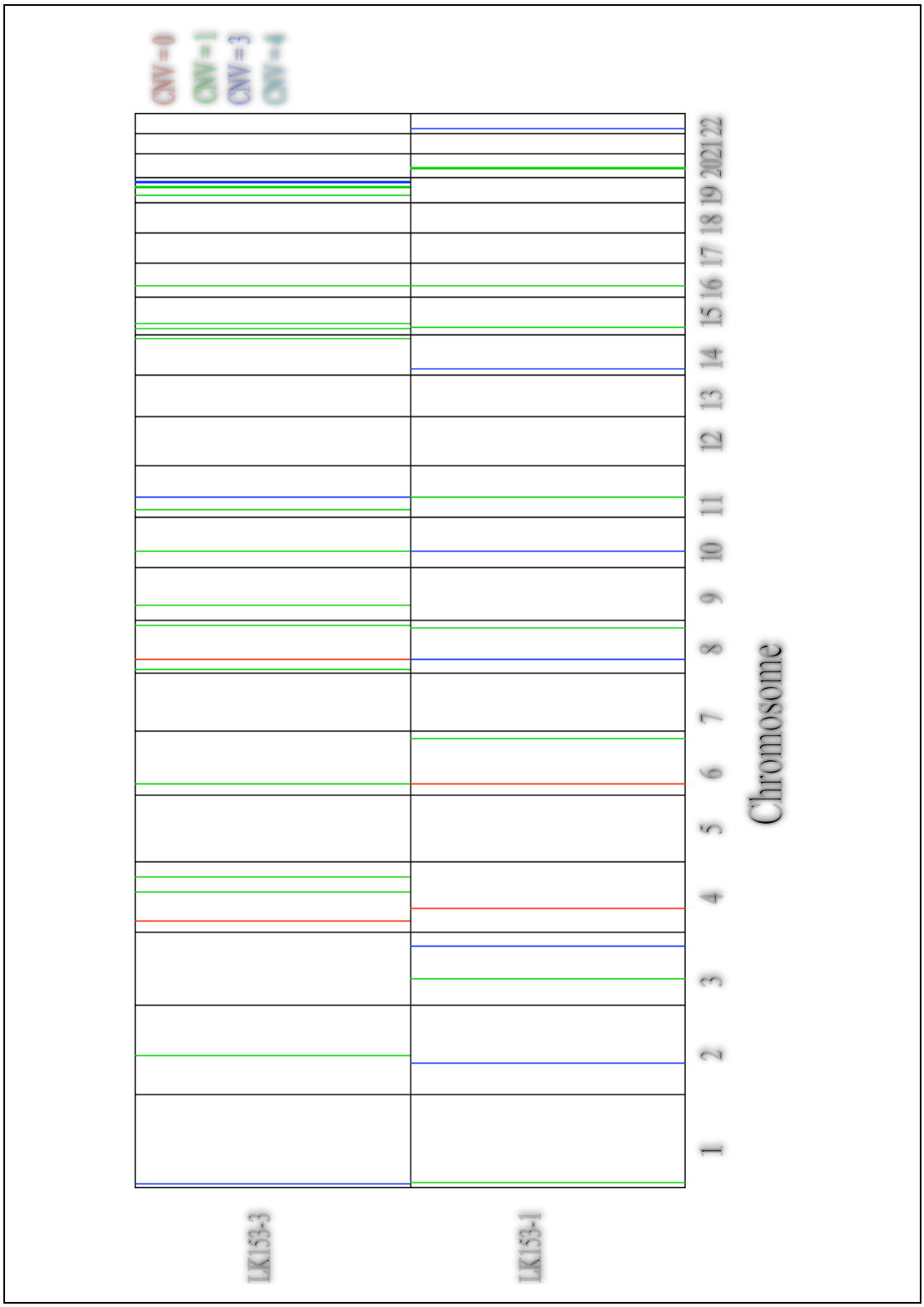
Figure 5.3.3.9.1. Shows the location and type of CNV for the LK132 family.



5.3.3.10. CNVs in the LK153 family

Two subjects from this family were included in this study. All (LK153-1 and LK153-3) were affected with a HM (see Table 5.3.1). The CNVs in this family are shown in figure 5.3.3.10.1.

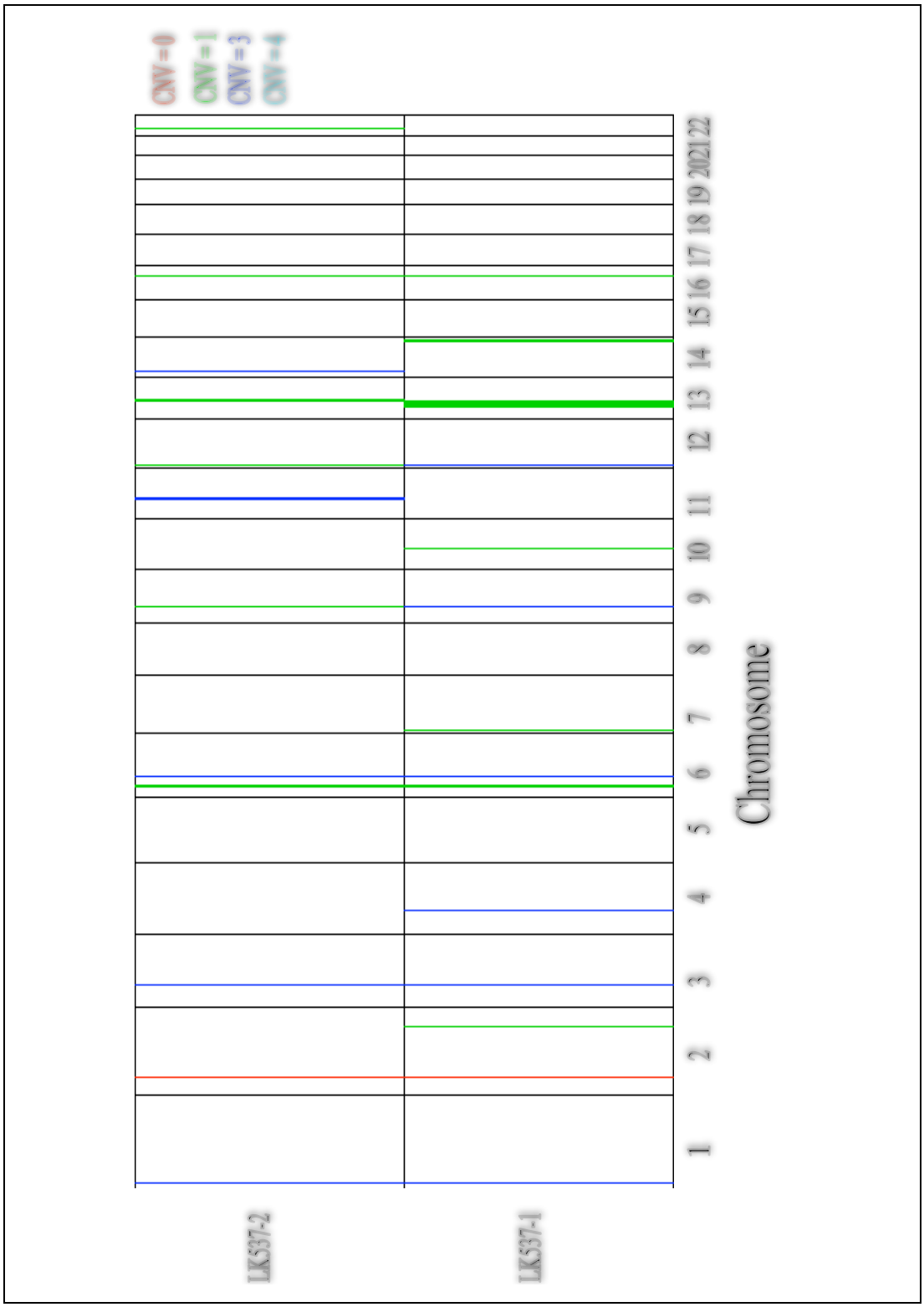
Figure 5.3.3.10.1: Shows the location and type of CNV for the LK153 family.



5.3.3.11. CNVs in the LK537 family

Two subjects from this family were included in this study. Both (LK537-1 and LK537-2) were affected with a HM (see Table 5.3.1). The CNVs in this family are shown in Figure 5.3.3.11.1.

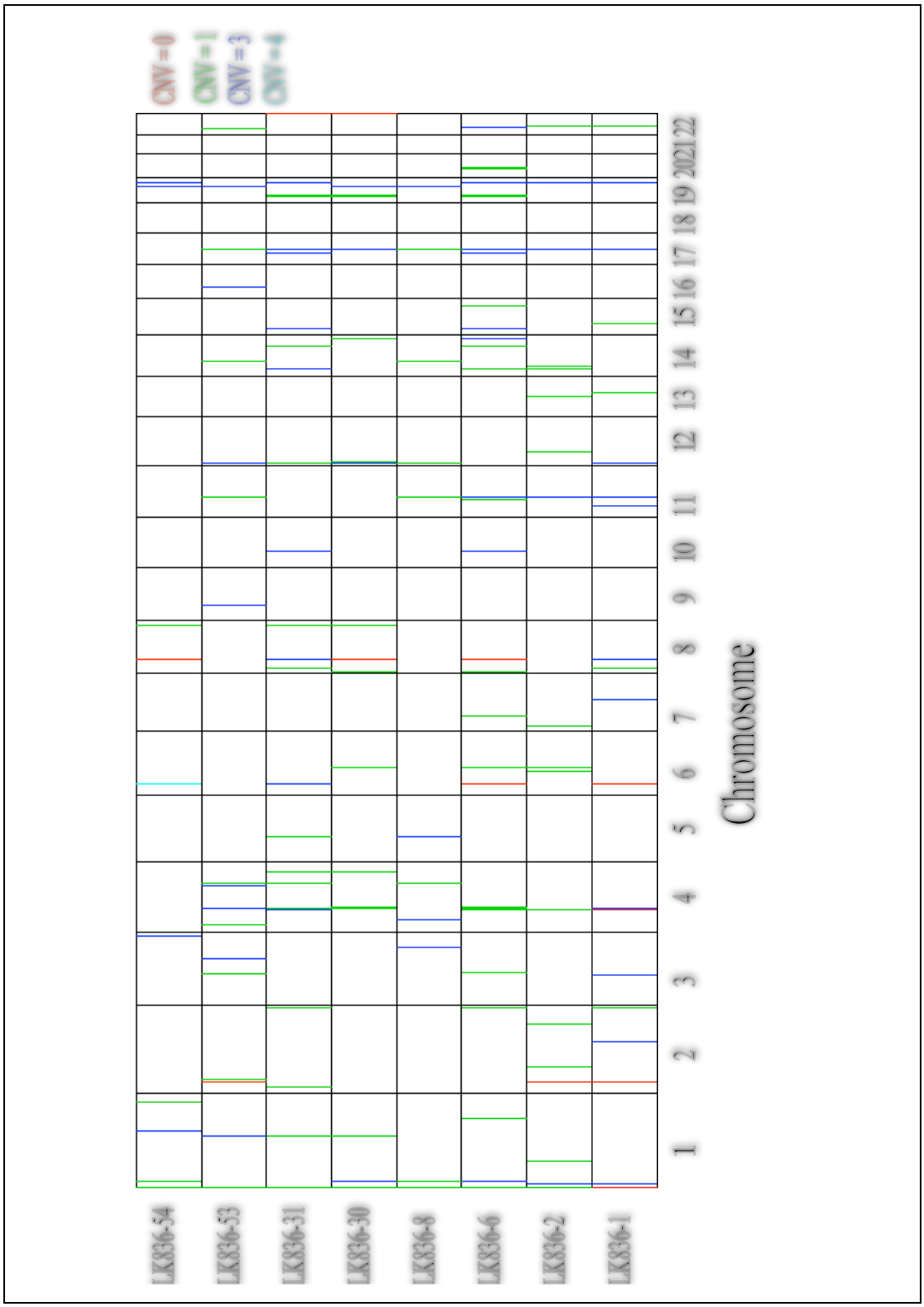
Figure 5.3.3.11.1: Shows the location and type of CNV for the LK537 family.



5.3.3.12. CNVs in the LK836 family

Eight subjects from this family were included in this study. Two (LK836-1 and LK836-2) were affected with a HM and the remaining six were all first-degree relatives of 3 affected members (see Table 5.3.1). The CNVs in this family are shown in Figure 5.3.3.12.1.

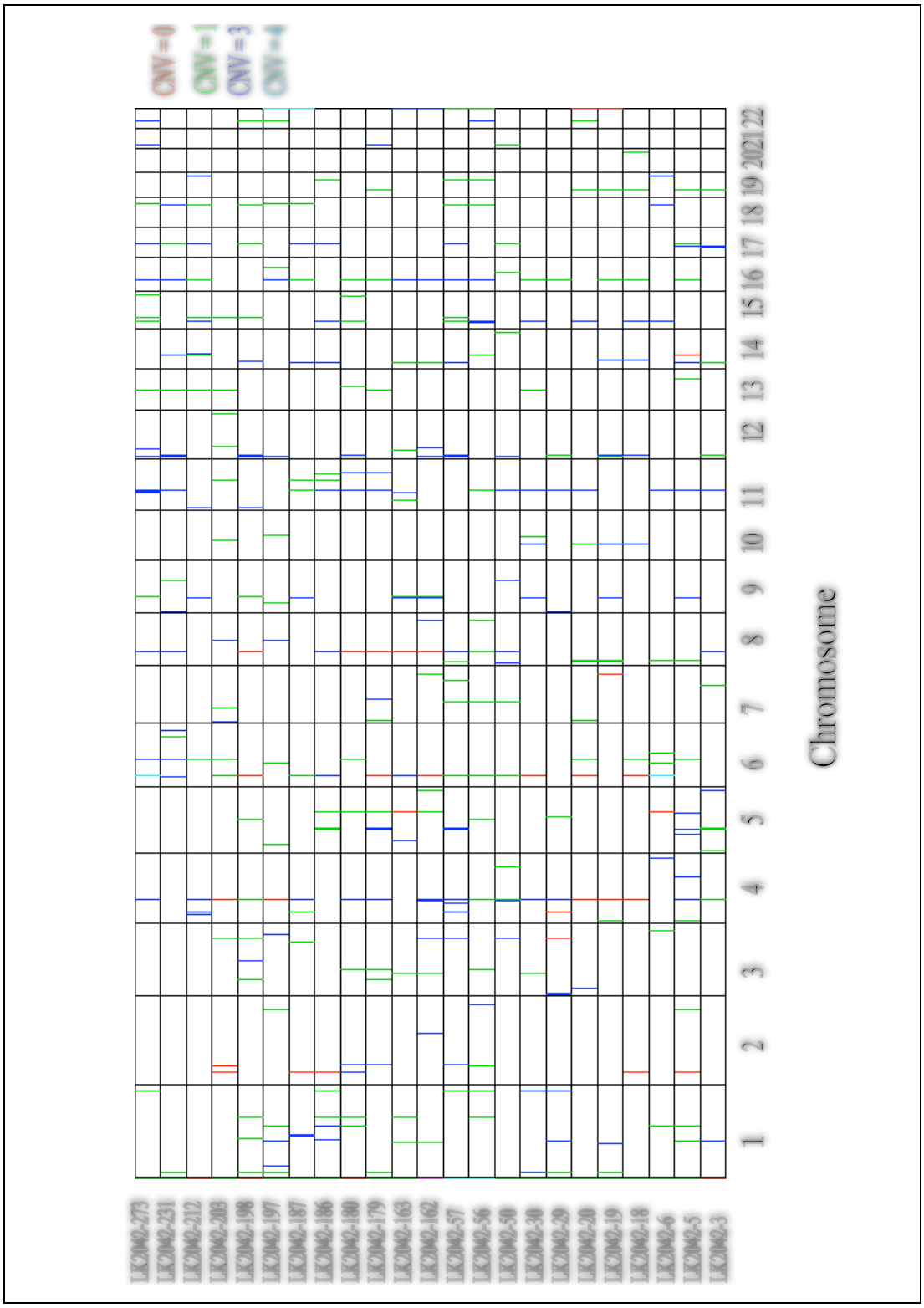
Figure 5.3.3.12.1: Shows the location and type of CNV for the LK836 family.



5.3.3.13. CNVs in the LK2042 family

Twenty-three subjects from this family were included in this study. Five subjects (LK2042-3, LK2042-5, LK2042-6, LK2042-129 and LK2042-231) were affected with a HM, the other eighteen were first-degrees of another ten affected members (see Table 5.3.1). The CNVs in this family are shown in Figure 5.3.3.13.1.

Figure 5.3.3.13.1: Shows the location and type of CNV for the LK2042 family.



5.3.3.14. CNVs shared in the tumour samples

LK16-1, LK537-1 and LK537-2 all had CLL and all had circulating disease when their DNA was extracted for this study. LK16-1, LK537-1 and LK537-2 have 3 CNVs in common; one on chromosome 13q14, a second on chromosome 12p13.31 and a third on chromosome 4q13. All three have a deletion of chromosome 13q14 (Figure 5.3.3.14.1) of different size, but they overlap. Chromosomal abnormalities at 12p13 are seen in about 1.4% of cases of CLL(198); however the CNV is not shared in the same number by all 3 (LK16-1 has 1 copy, LK537-1 has 3 copies and LK537-2 has 1 copy) and this region of 12p13 is a known CNV region. The third CNV that they share is shown in Figure 5.3.3.14.2(blue arrows). Table 5.3.3.14.1 lists all the people who share this duplication; This duplication has been confirmed by FISH (Figure 5.3.3.14.3) in LK16-15, with 3 regions for the G248P89388D9 fosmid probe (maps to chr4:69067720-69102943) and 2 signals for the G248P87131E4 fosmid probe (maps to Chr4: 69105512-69141702).

Figure 5.3.3.14.1: Copy number variations on chromosome 13.

The arrow on the right points to the large deletion on LK537-1 and the two arrows on the left point to LK16-1 and LK537-2.

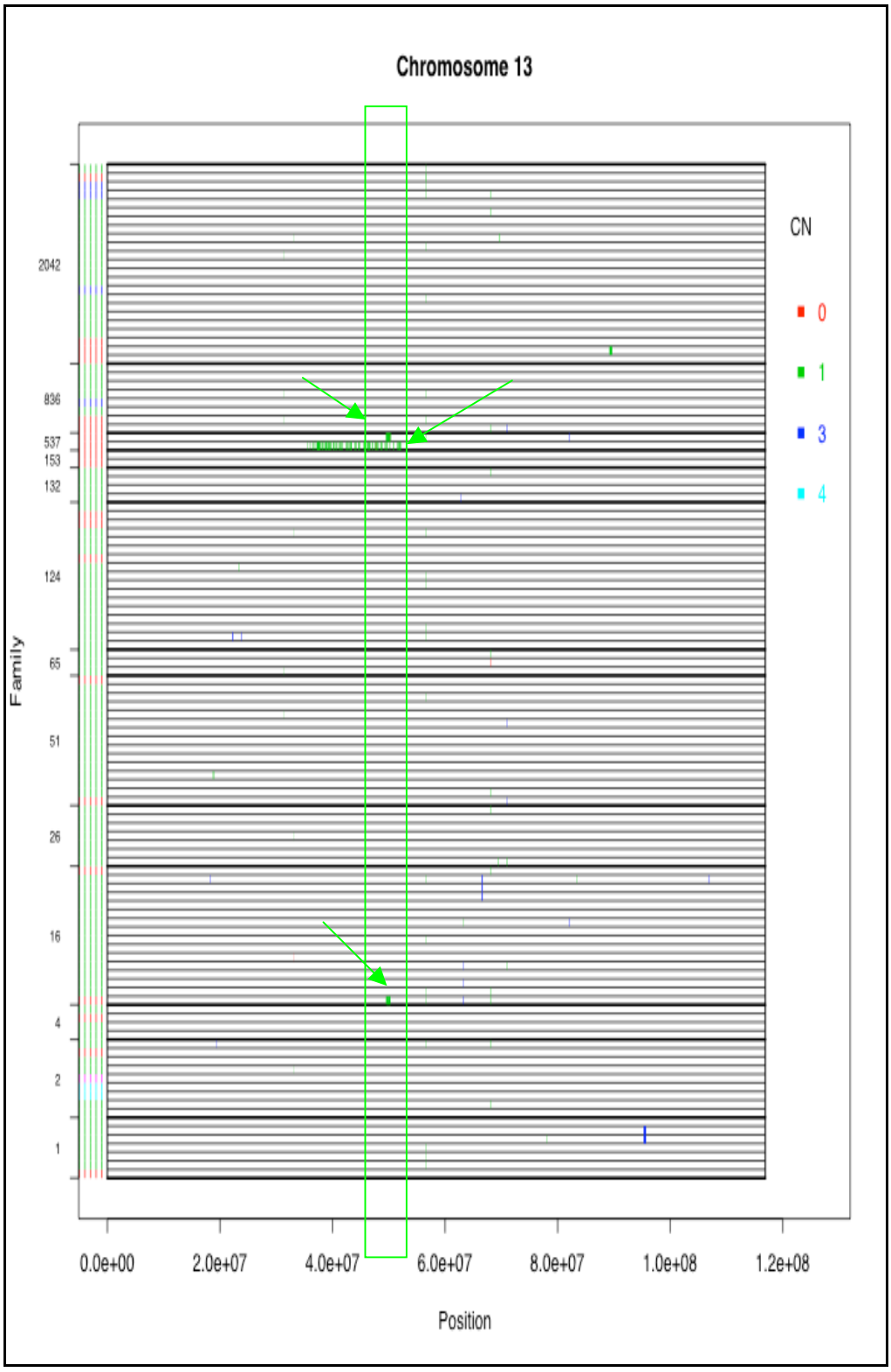
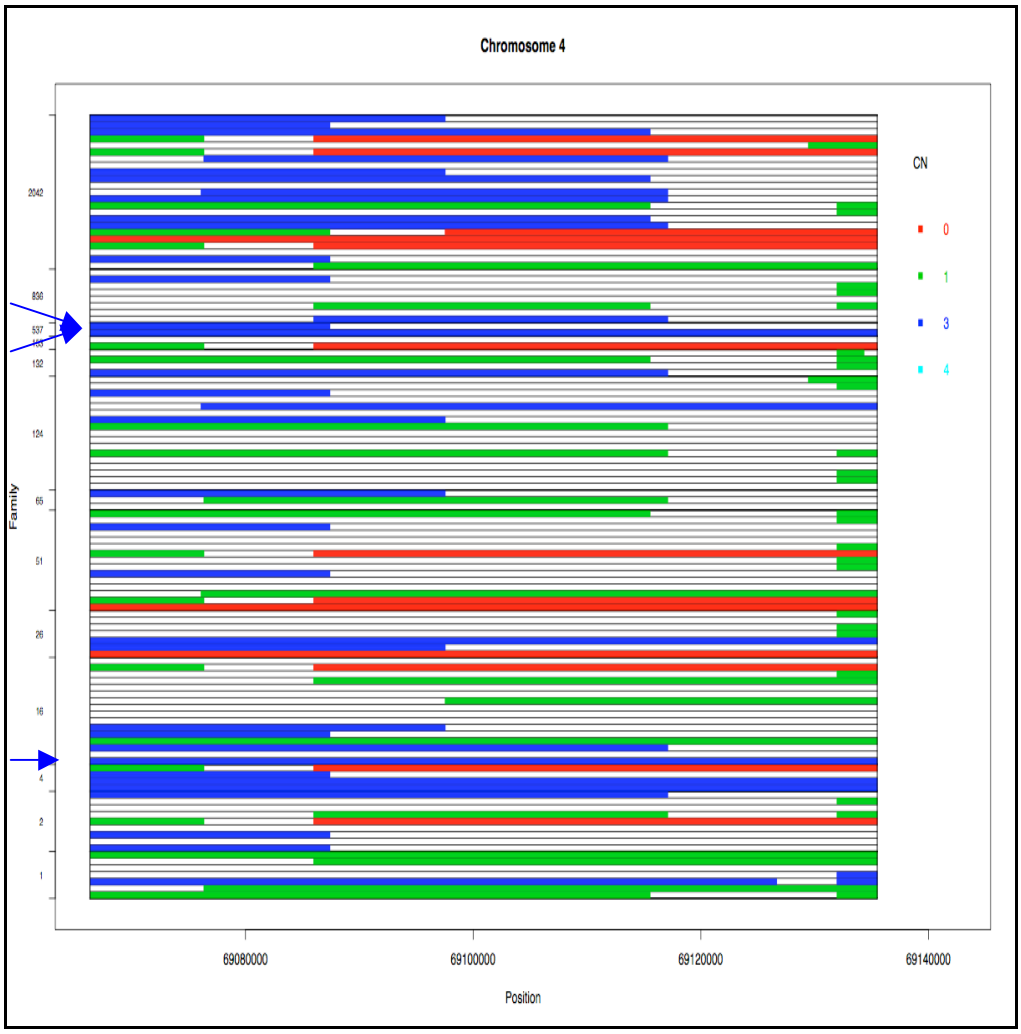


Figure 5.3.3.14.2: Shows the CNV on chromosome 4q13 in all 117 people.



Each row on the y-axis is an individual samples result, they are in numerical order from bottom to top (LK1-1, LK1-5…LK2042-231,LK2042-273). Blue is a CNV of 3 (duplication), green is a CNV of 1(heterozygous deletion), red is a CNV of 0 (homozygous deletion) and no colour is a CNV of 2 (normal).

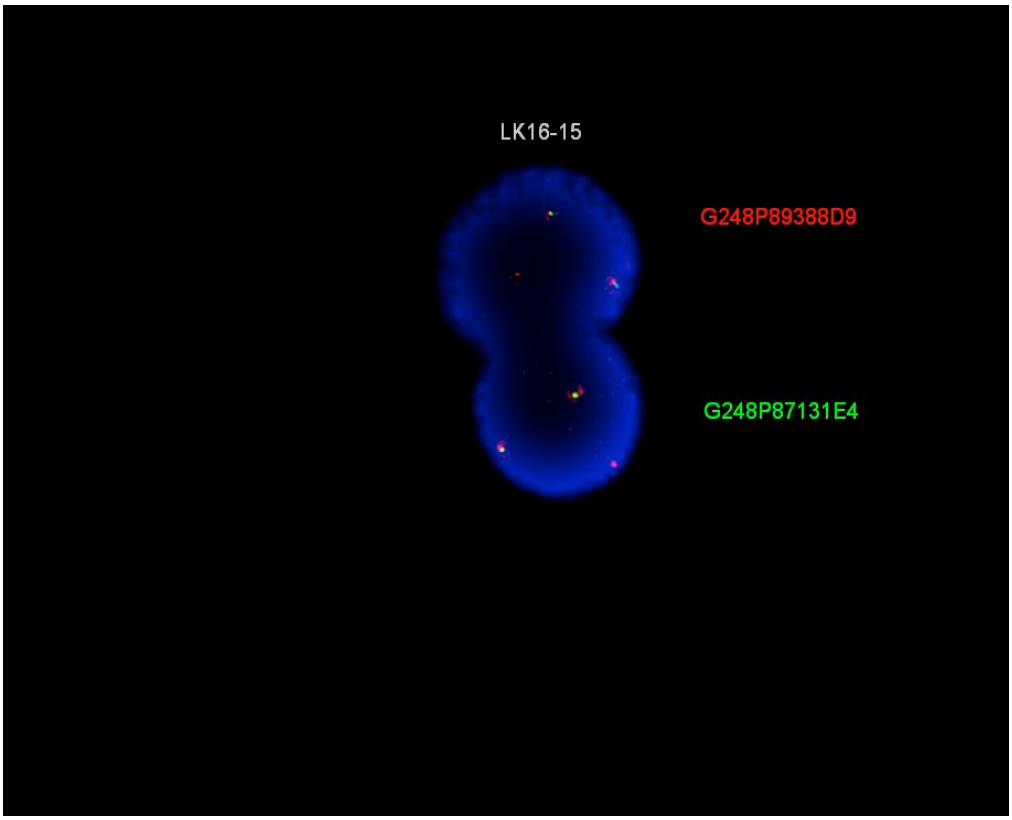
Table 5.3.3.14.1: All the people who have 3 copies of the chromosome 4q13 segment.

Position	ID	HM	Representing HM
chr4:69047457-69087412	LK0051-140		CML
chr4:69047457-69087412	LK2042-005	PBALL	
chr4:69064675-69087412	LK0002-022		MDS
chr4:69064675-69087412	LK2042-231	DLBCL	
chr4:69064675-69087412	LK0124-179	DLBCL	
chr4:69064675-69087412	LK0016-015	MBL*	
chr4:69064675-69087412	LK0051-075		HL
chr4:69064675-69087412	LK0002-011		NHL
chr4:69064675-69087412	LK0537-002	CLL	
chr4:69064675-69087412	LK0836-053		ALL
chr4:69064675-69097539	LK0065-034		MM
chr4:69064675-69097539	LK0124-117		AML
chr4:69064675-69097539	LK0016-016		CLL
chr4:69064675-69097539	LK2042-273		AML
chr4:69064675-69097539	LK2042-180		MM
chr4:69064675-69115543	LK2042-030		WM
chr4:69064675-69115543	LK2042-212		PBALL
chr4:69064675-69117090	LK2042-029		WM
chr4:69064675-69117090	LK2042-057		NHL

chr4:69064675-69126660	LK0001-062		RAEB
chr4:69064675-69135491	LK0537-001	CLL	
chr4:69064675-69163188	LK0004-007		DLBCL
chr4:69066337-69087412	LK0004-012	HL	
chr4:69066337-69097539	LK0026-009		PV
chr4:69066337-69115543	LK2042-129		MM
chr4:69066337-69117090	LK0016-005		CLL
chr4:69066337-69117090	LK0002-131		HCL
chr4:69066337-69117090	LK0132-016		MM
chr4:69066337-69135491	LK0016-001	CLL	
chr4:69066337-69163188	LK0026-010		PV
chr4:69066337-69163188	LK0004-008		DLBCL
chr4:69076097-69117090	LK2042-162		MM
chr4:69076097-69144894	LK0124-141		DLBCL
chr4:69076346-69117090	LK2042-187		ET
chr4:69085989-69117090	LK0836-001	DLBCL	
chr4:69047457-69087412	LK2042-005	PBALL	
chr4:69064675-69087412	LK2042-231	DLBCL	

**Developed MBL during the course of this study.*

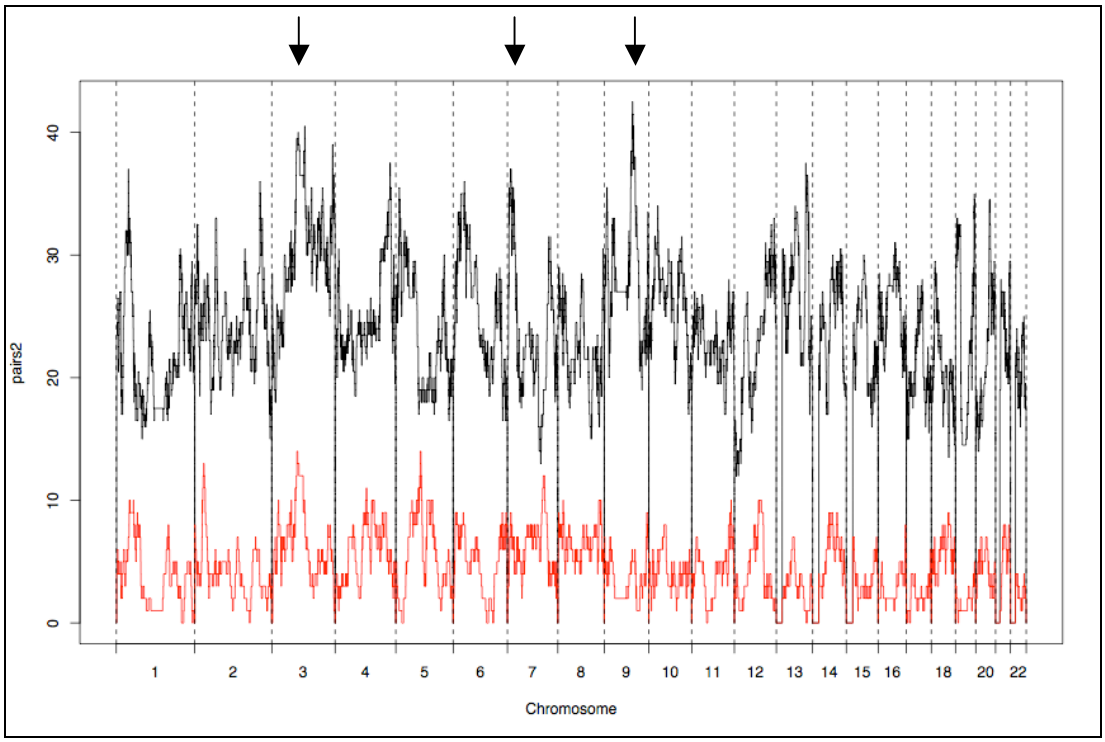
Figure 5.3.3.14.3: Two interphase cells from LK16-15, showing 3 red signals (duplication) and two green signals (normal) per cell.



5.3.4. PLINK analysis of IBD

The results of the analysis of the 13 families identifying regions of genotype sharing are shown in Figure 5.3.4.1. This paired analysis showed that the region where the most number of pairs are shared is on chromosome 9; the second highest region of sharing was on chromosome 3 and the third highest region was on chromosome 7.

Figure 5.3.4.1 Regions of chromosome sharing among all the families (black) compared to un-related Tasmanian controls (red).



5.3.5. Results of Sequencing of genes in these regions

5.3.5.1. Chromosome 9 peak (*DAPK1*)

The primers used in sequencing this gene are given in Appendix B. No mutation was found in this gene.

5.3.5.2. Chromosome 7 peak (*NDUF4*)

The primers used in sequencing this gene are given in Appendix B. No mutation was found in this gene.

5.3.6. The family based SNP association analysis

The top 5 SNPs that showed the most significant association in the families compared to a control dataset are listed in Table 5.3.6.1; and the reference allele frequencies of these 5 SNPs in Table 5.3.6.2. One of these top 5 SNPs (rs4314718) is near the region identified by the PLINK analysis on chromosome 9.

Table 5.3.6.1: Highest ranking SNPs.

SNP	Chr	Position (bp)	Consequence	Gene	p-value
rs1011313	6	15633432	INTRONIC	DTNBP1	0.000007743
rs978846	12	33128833	Gene desert - closest:	(PKP2)	0.000024000
rs156109	5	131626611	INTRONIC	SLC22A4	0.000034400
rs4314718	9	112013469	UPSTREAM	(C9orf152)	0.000041600
rs11489600	7	102735276	SPLICE_SITE,INTRONIC	PMPCB	0.000046300

Table 5.3.6.2: Top 5 SNP hits reference allele frequency based upon HapMap data.

SNP	Ref Allele	CEU	JPT	YRI
rs1011313	T	8.4%	13.4%	0.4%
rs978846	T	78.3%	25.6%	83.3%
rs156109	C	31.9%	15.7%	42.5%
rs4314718	A	67.3%	23.3%	71.7%
rs11489600	T	33.1%	65.1%	91.8%

5.4. *Discussion and Conclusions*

A genome-wide SNP array in 206 families by Sellick *et al.* 2007, identified three regions in which genetic mutations were located; but no specific disease causative mutation was identified. This study involved 206 families representing 550 affected people in those families(18). This results in a mean of 2.66 affected people per family. This is a relatively low number of affected in each family, compared to the 13 Tasmanian families where there is a mean of 10 people affected (2-27 people) and a mean of 6 (2-16) people per family represented in this SNP array. The likelihood of finding a mutation is increased with more densely affected families. The other difference between these families and the 13 Tasmanian families is that the affected people in the families studied by Sellick *et al.* 2007, all had CLL, whereas there is a range of different HMs in the 13 Tasmanian families. This difference, and hence the use of families that have different types of HMs has been postulated by Shpilberg *et*

al. 1994, as supportive evidence of the mutation being present in the pluripotent stem cell(19).

Three people had circulating tumour cells at the time of their DNA collection in this study. Therefore, the CNVs in these 3 patients might represent acquired CNVs in their tumour cells. A recent study found that acquired CNVs could be detected when as little as 30% of the starting material is from the tumour(172). All three people with circulating tumour cells (LK16-1, LK537-1, LK537-2) had CLL.

One person (LK537-1) had significantly more CNVs than anyone else genotyped (92 compared to a mean of 35 CNVs). The increased CNVs in this subject were all due to a large heterozygous deletion on chromosome 13q14. Figure 5.3.3.14.1 shows the CNVs on chromosome 13. LK537-2 and LK16-1 also show a heterozygous deletion within the region that LK537-1 has deleted. A deletion of chromosome 13q14 is a commonly seen acquired deletion in CLL (172) and this results in the deletion of two microRNA genes (miR15 and 16)(199). This was the most common deletion that was found in a study of acquired CNVs in CLL(192).

The CNV that these three people share on chromosome 4q13 is interesting. There is a gene in this region, UDP-glucuronosyltransferase (*UGT2B17)*. *UGT2B17* encodes an enzyme that catalyses the transfer of glucuronic acid from uridine diphosphoglucuronic acid to a variety of substrates including steroid hormones. This process, known as glucuronidation, is an intermediate step in the metabolism of steroids(200).

CNVs of *UGT2B17* have been reported to be associated with disease, in particular osteoporosis and prostate cancer (201, 202). Homozygous deletions of *UGT2B17* (a CNV of zero) have also been implicated in causing donor-recipient mismatch in graft

versus host disease(203). This gene has only ever been reported in the literature to have 0,1 or 2 copy numbers per person (201); however, in the 117 people in this study, several members had 3 copies of this region. In fact 32% of people had 3 copies of this region. This CNV analysis was repeated in another dataset of Australian subjects with uveitis and found that approximately 30% of these patients have 3 copies of this region (This analysis was done by Dr Patrick Danoy). FISH with fosmid probes mapped to this region were also used to confirm this duplication (Figure 5.3.3.14.3). It is interesting to note that in Caucasian populations this CNV seems to be commonly duplicated, while in Asian populations it is commonly deleted; which correlates with the incidence of CLL in these populations. Houlston *et al*. 2002, previously reported a CLL predisposition allele located at 4q11-q21(204); 4q13 is in the middle of this predisposition region. Therefore, an increased copy number of this region may be associated with an increased risk of CLL. It would be interesting to assess this in a larger number of CLL patients. There were 8 CLL affected individuals genotyped and at least 5 of them had 3 copies of this region (62.5%).

During the course of this study, LK16-15 developed MBL; therefore the CNVs that LK16-1 and LK16-15 had were compared to assess for a germline CNV segregating in this family. As they are related, it is not unexpected that they should share some CNVs as these are inheritable traits. It was found that they shared CNVs on chromosome 4 (64381774-64398956 & 69064675-69082412), chromosome 6 (56040470-56140734), chromosome 9 (43613173-43616717), chromosome 12 (9535285-9607393) and chromosome 13 (63231043-63277373). It is intriguing that

they share the duplication on chromosome 4q13 and this would support this being a predisposition allele.

PLINK analysis of IBD sharing across all the 13 Tasmanian families showed that the region of greatest sharing was on chromosome 9. It was interesting that this corresponded to a region seen in another CLL family(25), in which the only causative familial HM mutation (affecting the expression of *DAPK1*) was found. However this mutation and any other mutation in *DAPK1* was not found in the 13 Tasmanian families.

As only one gene (*NDUFA4*) was located in the chromosome 7 region of sharing, this was sequenced and no mutations were identified. The region of sharing on chromosome 3 contains many genes, several of which have been implicated in HMs. The results of their sequencing in the present study are shown in Chapter 7. Formal linkage analysis is also underway to confirm the PLINK analysis.

Analysis of SNPs that showed an increased association in these 13 Tasmanian families compared to controls identified several SNPS with a significant association in these 13 Tasmanian families. One region that has previously been identified in the literature as harbouring a predisposition allele (6p22.1)(18) is located very close to a SNP (rs1011313 6p22.3) identified in these 13 Tasmanian families. It is becoming clear in the literature that multiple genes and multiple mutations are going to be responsible for the heritable risk of HMs (18, 149).

In conclusion, the genome-wide SNP array has identified several different types of genetic loci (SNPs, CNVs and genetic loci) that require further follow up in another dataset of patients with HMs. It would be very interesting to follow up these results in patients from different ethnic backgrounds.

6. Chapter 6: Cell lines and gene expression

6.1. *Introduction*

Another approach used to identify mutations in families has been to assess the differential gene expression in affected individuals compared to normal controls(205, 206). In one CLL family, decreased expression of *DAPK1* was found in affected members compared to unaffected family members(25); and subsequent sequencing of this gene found a SNP up-stream of *DAPK1* (c.1-6531A>G) which resulted in increased binding of *HOXB7* causing decreased expression of *DAPK1*(25).

The first report of a gene expression array was published in 1996 (207). This first array was based on hybridization to small, high-density arrays containing thousands of synthetic oligonucleotides, which in turn are based upon sequence information and synthesized in situ using both photolithography and oligonucleotide chemistry(207). There are a variety of expression microarray platforms that are available for use today (spotted cDNA arrays, Affymetrix GeneChip® arrays, Agilent ink-jet arrays and Illumina long-oligonucleotide bead-based arrays)(208).

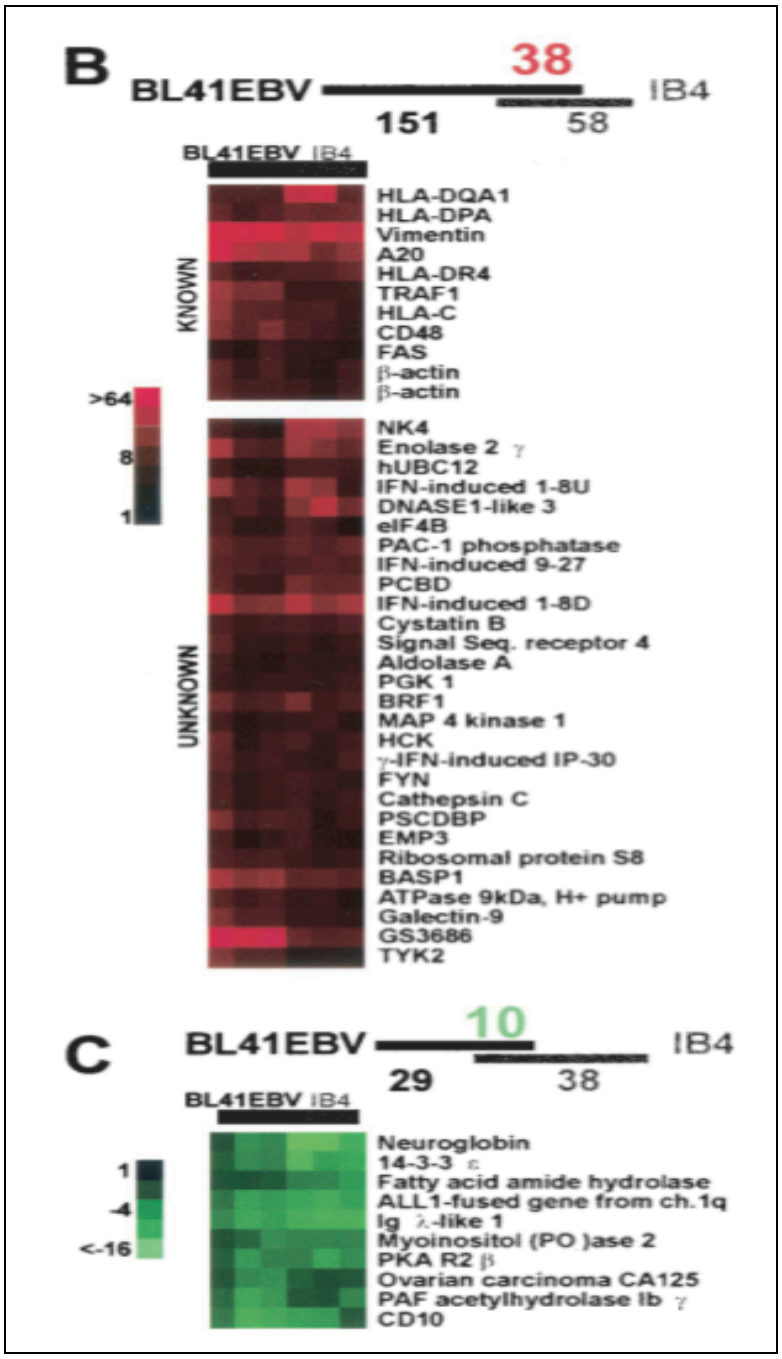
The Affymetrix GeneChip® Human Exon 1.0 ST array has been available for commercial use since 2006, and contains over six million probes targeting all annotated and predicted exons in the human genome(209). This array was chosen for this experiment because its new design allows gene expression profiling at both the exon level and gene level using a single array, with at least 30 distinct probes representing most genes(210).

Epstein Barr Virus (EBV) transformation of peripheral blood lymphocytes can be used to establish permanent lymphoblastic cell lines(211). This is an established

method and has the following advantages: firstly peripheral blood lymphocytes are easy to obtain from patients; secondly EBV transformed cell lines have been shown to exhibit chromosomal stability; thirdly RNA/DNA can be readily extracted from the cell lines; and finally they are a continuous/permanent resource (211).

EBV infects peripheral blood B lymphocytes and transforms them by altering cellular gene expression through the expression of viral *EBNA2*, *EBNALP*, *EBNA3A*, *EBNA3C*, *EBNA1*, and *LMP1*(212). When a person is infected with EBV, a significant percentage of their B lymphocytes are latently infected and express virus-encoded proteins; a process called latency III infection(212). The gene expression profile of latency III lymphocytes is also characteristic of the gene expression in EBV transformed lymphoblastoid cell lines(212). The genes that show an altered expression in latency III EBV infection are listed in Figure 6.1.1(212). Whilst EBV transformation will affect the gene expression profile, in this analysis gene expression differences between the homozygous, compared to the heterozygous, compared to normal controls; will be assessed and EBV transformation should affect all samples equally.

Figure 6.1.1: The reported effect of changes in gene expression in latency III EBV infection (212).



(B) Intersection of genes induced by EBV twofold or greater in both BL41EBV and IB4

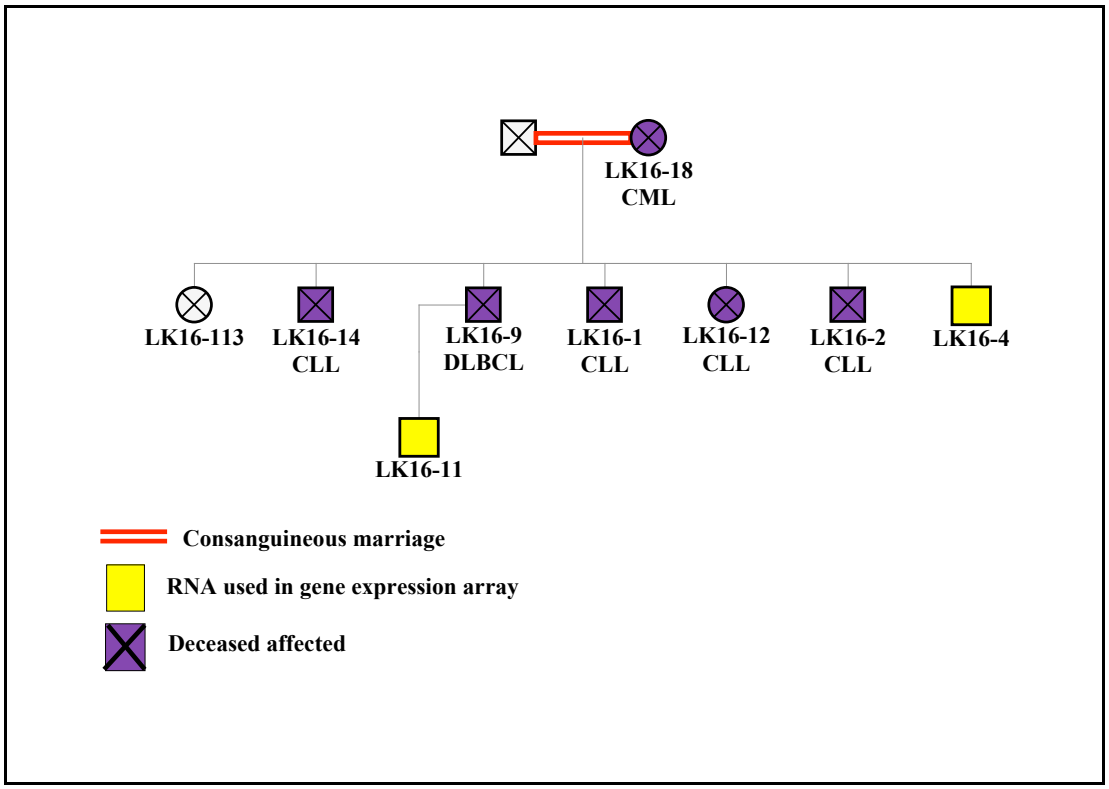
(C) Intersection of genes repressed by EBV twofold or greater in both BL41EBV and IB4

The hypothesis in the LK16 family is that a homozygous incompletely penetrant gene mutation is segregating in this family. It is proposed that examination of the

differential gene expression in the family may provide altered gene expression data that could be compared with the linkage analysis. The LK16 family is described in Chapter 3 and Chapter 4. A region on chromosome 15 between 30919517 and 37872929 base positions was found in this family where the affected siblings (LK16-1, LK16-9, LK16-12, LK16-14 and LK16-2) were homozygous, with four of them being homozygous from 30919517 to 57481797 base positions.

Ideally, the differential gene expression was to be assessed in the LK16 family (Figure 6.1.2): in multiple affected members who are homozygous for the chromosome 15 region; a member heterozygous for the chromosome 15 region; and in unrelated controls. However, this proved impossible because the last remaining affected sibling (LK16-1) was now deceased. However his last surviving sibling (LK16-4) was still alive and homozygous for the whole region on chromosome 15. A nephew (LK16-11) was used as the heterozygous carrier.

Figure 6.1.2: Pedigree of the LK16 family showing members represented in the gene expression and their relationship to the index case (LK16-1).



6.2. *Materials and Methods*

6.2.1. Cell lines

Blood was collected in Ethylenediaminetetraacetic acid (EDTA) from LK16-4, LK16-11 and 2 unrelated males matched for age. Cell lines were made by standard methods using EBV to transform B lymphocytes in combination with cyclosporine (Sigma, Australia) to suppress a T lymphocyte response as described in chapter 2.11. All samples were treated in the same manner.

6.2.2. Molecular Cytogenetics

Lithium heparin peripheral blood samples were obtained form LK16-4. Chromosome preparations were obtained as described in Chapter 2.9 and results reported as per ISCN(161). FISH probes from Vysis (Abbott Molecular, Australia) for *IGH@*, 13q14, 13q34, CEP(12), *TP53* and *ATM* were used as described in Chapter 2.10.

6.2.3. Statistical analysis

RNA from LK16-4, LK16-11, and two normal controls matched for age and gender were sent to the Australian Genome Research Facility for genome-wide Affymetrix GeneChip Human Exon 1.0 ST array. The analysis of the gene expression data was performed by Professor Gordan Smyth and Yifang Hu of the Walter and Eliza Hall Institute using R (The R Foundation for Statistical Computing). The analysis was undertaken in 3 parts: firstly, analysis at the gene level for genes located on chromosome 15:30919517 to 37872929; secondly, analysis at the exon level for all exons located on chromosome 15:30919517 to 37872929; and thirdly, for all genes on a genome-wide analysis. The levels of gene expression were compared in the homozygous to the heterozygous compared to normal controls.

6.2.4. PCR primers

Primers for genes and differentially expressed exons were designed and optimised as per Chapter 2.8. Sequencing of genes was done on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems) as described in Chapter 2.8.

6.3. *Results*

6.3.1. Molecular cytogenetics

This was undertaken to confirm that LK16-4 did not have the known common acquired mutations that occur in CLL. Deletions of chromosome 13q14, *ATM*, *TP53* and trisomy 12 are common cytogenetic abnormalities seen in CLL(213). FISH using the commercially available probes from Vysis (Abbott Molecular, Australia) for the common abnormalities found in CLL (Probe set 1:LSI D13S319/13q34/CEP 12: Probe set 2: LSI *ATM/TP53*) were normal in 200 metaphase and interphase cells examined. The FISH images are provided in Figure 6.3.1.1 for probe set 1 and Figure 6.3.1.2 for probe set 2.

Figure 6.3.1.1: FISH image of LK16-4, using the Vysis probe set 1 for CLL*, showing a normal metaphase cell (2 signals for 13q14 (red), 2 signals for 13q34(aqua) and 2 signals for centromeric region of chromosome 12(green)).*

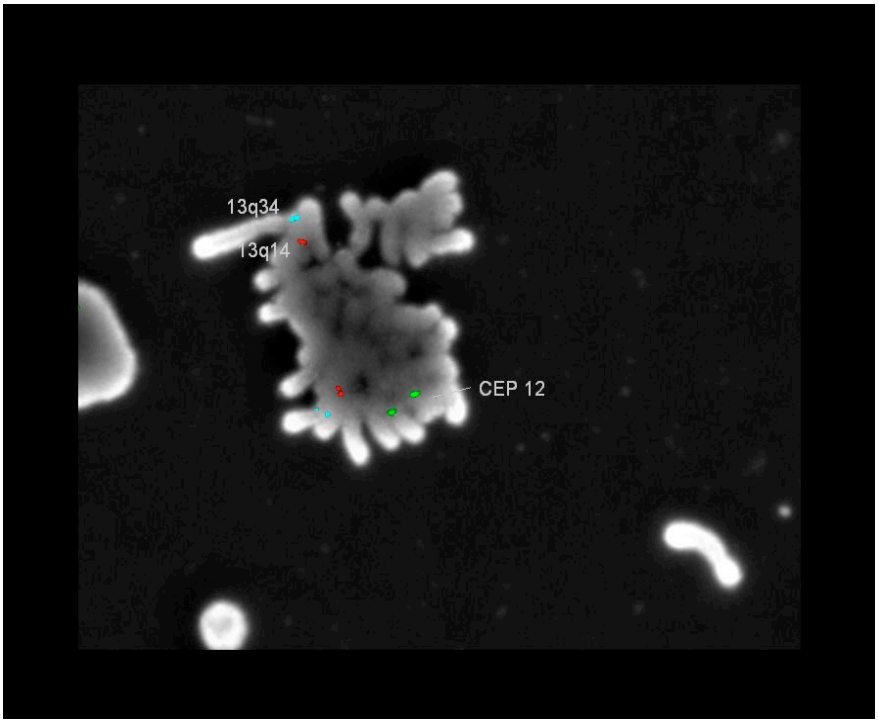
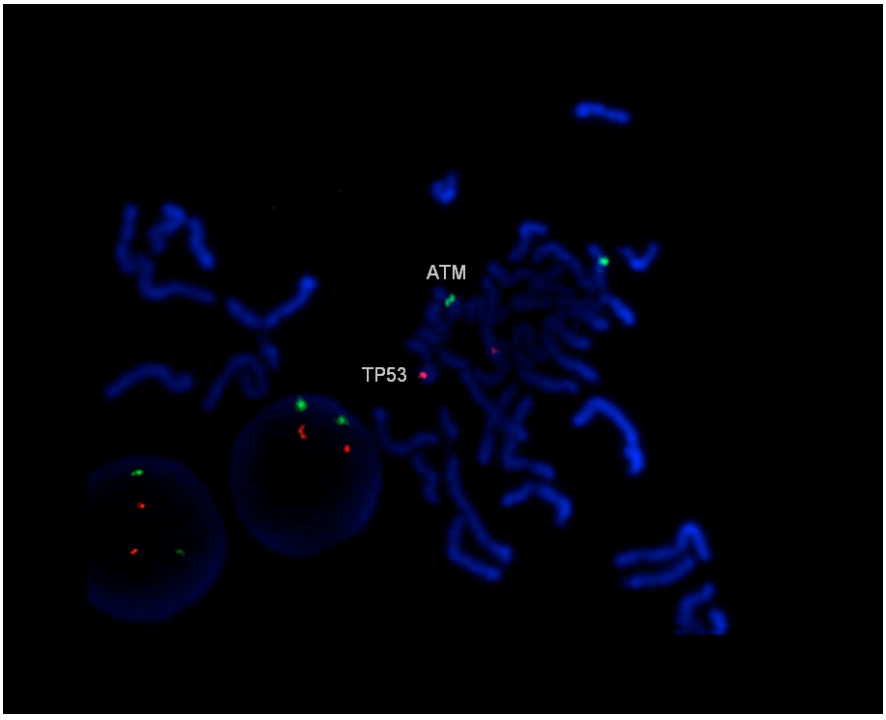


Figure 6.3.1.2: FISH image of LK16-4 using Vysis probe set 2 for CLL*, showing normal interphase and metaphase cell (2 signals for ATM(green) and 2 signals for TP53(red)).*



6.3.2. Statistical analysis of differential expression of individual exons and genes on chromosome 15:30919517-37872929 and of the all genes.

When the region of interest on chromosome 15 (Chr15: 30919517-37872929) was analysed at the gene level, the most differentially expressed gene was PGBD4 Statistical analysis of the differential individual exon expression for this same region found that several exons were differentially expressed. These exons and how they were differentially expressed are listed in Table 6.3.2.1.

Table 6.3.2.1: Exons within the region of interest that were differentially expressed

Gene	Position	Expression	P value
C15orf24	32175402-32175490	Increased	p=0.0269
PGBD4	32182228-32183724	**Decreased**	p=0.0286
TMEM85	32309295-32309320	Increased	p=0.0077
LPCAT4	32441777-32441814	**Decreased**	p=0.0447
GOLGA8A	32460866-32460901	Increased	p=0.0026
GOLGA8A	32466152-32466181	Increased	p=0.0155
GOLGA8A	32467345-32467400	Increased	p=0.0433
ACTC1	32871902-32871926	Increased	p=0.0217
AQR	32969743-32969809	Increased	p=0.0140
AQR	32997799-32997871	Increased	p=0.0182
C15orf41	34771623-34771668	Increased	p=0.0570

TMCO5	36022837-36022861	Increased	p=0.0121
FAM98B	36553005-36553029	Increased	p=0.0270
RASGRP1	36579596-36579630	Increased	p=0.0291
RASGRP1	36595690-36595764	Increased	p=0.0136

Analysis at a genome-wide level found that the most differentially expressed gene was RNF111 (this showed decreased expression, $p=1.9 \times 10^{-5}$). This gene is located on chromosome 15: 57067157 to 57176545.

6.3.3. Sequencing of genes and exons identified

Exons differentially expressed

The exons that were identified were sequenced as well as 100 base pairs on either side of the exon. The primers used to sequence these regions are listed in Appendix B. No mutations were identified in the coding regions or any splice site variations.

PGBD4

This gene is located on chromosome 15 at 32181566 to 32183883 base positions. PGBD4 stands for piggBac transposable element derived 4. It has only one coding exon. The primers used in the sequencing of this gene are given in Appendix B. No known or novel mutations were found in the sequencing of this gene or 500bp upstream or downstream of this gene in the LK16 family. Interestingly LK16-1 was AA for a SNP (RS1436923), which is located 1 base pair position upstream of this gene, the reference base for this position is C. However, the allele frequency for the A allele is 92% in the CEU population.

RNF111

The ring finger protein III (*RNFIII*) gene is located on chromosome 15 at base position 57067157 to 57176545. According to UCSC genes, it has 8 isoforms and a total of 15 coding exons. The primers used in sequencing this gene are given in Appendix B. Sequencing of RNF111 found a coding mutation in exon 2 that results in an amino acid change at the 9th position, asparagine to lysine. LK16-1 was homozygous for this mutation. The allele frequency of the "GG" genotype for this SNP (RS289964) is 18.3% in the CEU population. The GG genotype is not seen in African-American or Asian populations. The results of the genotype for this SNP in 12 unrelated CLL patients and 22 normal controls are provided in Table 6.3.3.1. There were an increased number of CLL patients with a homozygous "GG" at this SNP compared to unaffected controls (15.3% compared to 9.1%, however this is not statistically significant increase).

Table 6.3.3.1: Genotype frequency data for RS2899642 SNP in sporadic CLL patients and normal controls

Genotype	CLL patients	Controls
CC	38.5%	50%
CG	46.2%	40.9%
GG	15.3%	9.1%

6.4. *Discussion and Conclusions*

Gene expression profiling has been used in many types of HMs. This has mainly been used for assessing the gene expression profile of the HM cells: which can then be used to classify and predict outcome in ALL(214); MLL positive leukaemias(215); and mature B cell neoplasms(216). Gene expression differences have been used successfully in one other CLL family(25) reported by Raval *et al.* 2007, and this information helped to identify a causative predisposition mutation. This family, when first published, consisted of a father and four sons affected with CLL(170). Interestingly the people affected with CLL in this family also had an acquired deletion of 13q14 in their CLL cells(170), just like LK16-1. This pedigree has been extended and a third generation is now affected with CLL(25). Linkage analysis in this family was to chromosome 9 between markers D9S175 to D9S1776. This region included 11 predicted genes and 3 known genes, one of which was *DAPK1*(25). Another study had shown an increased methylation of the *DAPK1* promoter in CLL cells(217), resulting in decreased expression of this gene.

Therefore, they assessed *DAPK1* expression by RT-PCR and this was found to be decreased in this family(25). Sequencing of this gene region found a SNP (c.1-6531 A>G) in the CLL samples, but not in normal controls. This SNP is located at a *HOXB7* binding site and this SNP increases the affinity of *HOXB7* binding which results in a decreased expression of *DAPK1*(25).

DAPK1 expression was reduced by approximately 25% in family members who are carriers of this SNP; however, a further reduction in the *DAPK1* expression levels occurred in the CLL cells due to acquired promoter methylation(25), these patients also had an acquired deletion of chromosome 13q14 in their CLL cells. This finding supports the hypothesis that further mutations have to occur in addition to the germline mutation in these families for a HM to develop. Interestingly LK16-1 also had a deletion of chromosome 13q14 in his CLL cells; however this is a common deletion in CLL(175) and results in down-regulation of two micro-RNA genes (miR15a and miR16-1). No conventional or molecular cytogenetic abnormality was found in LK16-4, the homozygous surrogate for LK16-1, in particular FISH did not identify a deletion of chromosome 13q14. In mice RNF111 has been reported to have a target site for miR15a(218).

EBV transformation of lymphocytes to make lymphoblastoid cell lines has been a standard technique employed for several decades(211). The gene expression changes of the EBV transformation process have been studied, and it has been found to result in similar changes to gene expression to that seen in cells that have a latent III EBV infection(212). It has been reported that only 38 genes have significantly altered expression as a result of EBV infection(212). None of these genes were identified in the LK16 family. None the less, it is a possibility that one of these genes was the

responsible gene in the LK16 family, and that the EBV process "normalised" the expression. However none of these 38 genes were located in the chromosome 15 region of interest.

The most differentially expressed gene (p=0.031) in the region of interest was piggyBac transposable element derived 4 (*PGBD4*). *PGBD4* belongs to the piggyBac transposon family of genes(219) and there are 5 in the human genome(219). *PGBD4* comprises a single exon and appears to be derived from the MER75 transposon(220). Transposons are segments of DNA that can move around to different positions in the genome of a single cell, and by doing so, they may cause mutations or increase/decrease the amount of DNA in the genome and are sometimes referred to as "jumping genes"(219). A SNP (RS1436923) that is 1 base pair upstream of *PGBD4* was found to be the non-reference allele in the LK16 family; however the non-reference allele frequency is 92.5% in people of European ancestry and of a similar frequency in other ethnic groups.

The most differentially expressed gene in the whole genome was *RNFIII*. *RNFIII* is located on chromosome 15 at base pair position 57067157 to 57176545. This is outside the region of interest; however LK16-1 (index case) and his brother (LK16-4), whose RNA was used in the gene expression array are homozygous from base pair position 30301633 to 57481797 on chromosome 15; which includes *RNF111*. The region of interest is narrowed from 30919517 to 37872929 because of LK16-15 whom only carries the haplotype to chr15: 37872929. This means that his mother (LK16-12), who is one of the affected siblings can only be homozygous from chromosome 15 base position 30919517 to 37872929 and is heterozygous from base

position 37872930 to 57481797 as her other son (LK16-16) is a heterozygous carrier of the full region of interest (30919517 to 57481797).

It has been reported that the leukaemic cells in CLL accumulate from deficiencies in apoptosis rather than by acute proliferation(221). *RNFIII* encodes for a protein that contains a RING finger domain, which is known to be involved in protein-protein and protein-DNA interactions(222). The mouse version of this gene (arkadia) has been shown to interact with transforming growth factor beta-like factor nodal and act as a modulator of the nodal signalling cascade(222). Transforming growth factor-Beta (*TGF-ß*) and *TGF-ß* related factors induce apoptosis in a many tissue types(223). It is known that *TGF-ß* appears to function in CLL as a negative regulator of B lymphocytes(224), and that *TGF-ß* induces the expression of death-associated protein kinase (DAP-kinase) as an immediate response in cells that undergo apoptosis in response to *TGF-ß* (223). It has been found in CLL that loss of responsiveness to *TGF-ß* and a decrease in *TGF-ß* receptor expression in CLL cells might provide a selective advantage to CLL cells(224, 225). It is also known that *DAPK1* expression is down regulated in CLL(25). Therefore it is plausible that *RNFIII* may play a role in CLL in this family, especially as it has been reported that genetic variants in apoptosis related genes are associated with the risk of CLL(221).

DNA sequencing of *RNF111* led to the finding of a mutation in exon 2. This is a known mutation of the gene (RS2899642) and results in an amino acid change from asparagine to lysine at the 9th amino acid position. LK16-1 was homozygous for GG at this SNP. This missense mutation is considered a normal variant. The reference allele for this SNP is C. The interesting anomaly about this SNP is its allele frequency in different populations. According to HapMap data the G allele frequency

in the CEU population is 45%, but in Japanese or African populations its frequency is only 10% and 0% respectively (http://hapmap.ncbi.nlm.nih.gov/cgi-perl/gbrowse/hapmap28_B36/#search). This G allele frequency also correlates with the incidence of CLL in different populations; as it is the most common leukaemia in western countries, but rare in Africa and Asia(35). It is also known that people of an Asian background keep their low risk of CLL if they move to a western country(35). This SNP was assessed in 12 unrelated people with CLL from Tasmania and the frequency of the G allele was 38.5%, which is the expected allele frequency. The allele frequency of this SNP was also assessed in 22 unrelated Tasmanians who did not have a HM, the G allele frequency was 29.5%. It would be intriguing to assess this SNP in people who have CLL and who are of Asian ethnicity.

In the literature, it has been suggested "that common genetic variants with modest effects on disease susceptibility may be related to risk"(221). Therefore, the genetic causes for familial CLL and other HMs may arise as a result of the complex interactions of multiple genetic loci. Many of the common loci identified as significantly associated with CLL are common. Therefore RS2899642 may additionally contribute to the risk of CLL. However, this needs to be validated in another dataset. Due to its different allele frequency in different populations, it would be very interesting to assess the G allele in CLL patients who are of Asian descent.

7. Chapter 7: Whole genome sequencing

7.1. Introduction

The technical ability to be able to do a whole genome sequence (WGS) of an individual person has dramatically advanced in recent years(226). The existence of chromosomes has been known since Karl Wilhelm von Nageli observed them under a microscope in 1842. Mitosis was discovered by Walther Flemming in 1882 and in 1912 Hans von Winiwarter declared the diploid number of chromosomes in humans to be 48(227). However this was revised to 46 in 1956 by Tjio and Leven(228). The molecular structure of DNA was first reported by Watson and Crick in 1953(229). Sequencing of DNA was achieved in 1977(230) and became known as Sanger sequencing after the author of this paper. Soon after this, the first gene sequence was published(231). The next major advancement in sequencing was the use of fluorescent dyes attached to the nucleotides, which then lead to the first automated DNA sequencer.

The sequence of the human genome was first published in October 2000(232). This was actually a combination of five people's genomes and took 3 to 4 years to complete and it cost approximately $300 million(157). In 2007 a single person's genome (J. Craig Venter's) was sequenced using Sanger dideoxy technology(226). Also in 2007, James Watson's genome sequence was published; however this genome was sequenced by next generation technology and was completed in 2 months(159).

With this new technology, originally called next generation sequencing, but now often referred to as second generation sequencing(156), it is possible to sequence all

the known exons (exomes), expressed genes (transcriptomes) or the complete genomes of people or their tumour cells. This allows for a comprehensive genome-based classification of cancer which should ultimately lead to better therapy for specific mutations(156); as well as aiding in the diagnosis of constitutional genetic disorders.

A paper by Ley *et al*. 2008 reported the whole genome sequencing of a cytogenetically normal AML genome(107). Approximately 40 to 50% of cases of AML have normal cytogenetics at diagnosis(233); however this does not imply that genetic mutations have not occurred, but are instead below the resolution of conventional cytogenetics(233). Many of these normal cytogenetic AML are known to have specific genetic mutations and this is reflected in the current WHO classification system of AML, which now includes a specific sub-type that have recurrent gene mutations(1). The sequencing of the cytogenetically normal AML genome identified two mutations in genes known to be commonly mutated in AML and an additional eight mutations in eight genes that were previously not thought to be involved in AML(107). The functionality of these mutations is still being assessed, however they would not have been found by any other approach. In the literature there are nine key whole genome sequencing studies to date(156); two of which are on HMs, both AML genomes.

The 1000 genome project aims to characterise the variation in the human genome so that correlation between phenotype and genotype can be made(234). To date they have published a pilot study comparing different types of platforms for WGS. They undertook 3 projects; firstly low coverage WGS of 179 individuals from four different populations, secondly high coverage WGS of two parent offspring trios, and

thirdly exon sequencing of 697 individuals from seven different populations(234). This project claims to have catalogued over 95% of the variants found in any individual's WGS(234). They have found that on average each person's WGS shows 250-300 loss of function variants and 50-100 variants previously implicated in inherited disorders(234).

LK16-1 was selected for WGS as he was the index case in a seminal family and a region of interest had been narrowed down to chromosome 15 (Chapter 4). The DNA obtained from LK16-1 was from peripheral blood lymphocytes; however at the time of the DNA collection his white cell count (WCC) was 60 x 10^9/L and comprised mainly lymphocytes, in keeping with his diagnosis of CLL. Therefore the WGS would be of tumour cells and some of the mutations would represent acquired changes in his CLL cells rather than a germline mutation. However as the hypothesis in this family is that the affected siblings inherited one mutation from each parent, a homozygous mutation is implicated. WGS study of cytogenetically normal AML cells had found that the acquired mutations were all heterozygous(160).

7.1. Methods and Materials

DNA from LK16-1 was chosen for whole genome sequencing. The DNA was extracted as described in Chapter 2.2. Sequencing was performed by Knome (Cambridge, MA, USA) on Illumina platform using pair end reads. The sequence reads were aligned to their reference genomes and to version Mar. 2006 (NCBI36/hg18) of the human genome and this was also undertaken by Knome (Cambridge, MA, USA).

7.2. *Results*

7.2.1. Whole Genome Sequence quality

WGS was undertaken on LK16-1 using DNA derived from his CLL cells. The mean sequencing depth was 37.45 reads per base with an average read length of 107.91 (See Table 7.2.1.1). Sequencing of the X and Y chromosomes showed no heterozygous calls, consistent with a male gender for LK16-1.

Table 7.2.1.1: Summary of sequence coverage and quality

Sequence coverage/quality measures	*Number*
Effective read length (bp)	107.91
Mean sequencing depth	37.45
Mismatches per site per read	0.008
Sites covered (of 3080436051)	2776947377
Sites called heterozygous	2217344
X-Specific sites call	154913754
X-Specific sites call heterozygous	0
Y-Specific sites call	25652954
Y-Specific sites call heterozygous	0
Autosomal sites call	2677446101
Autosomal sites call heterozygous	1975513

7.2.2. Mutations found in the genome of LK16-1

A summary of the type and number of mutations/polymorphisms found in the genome is shown in Table 7.2.2.1. There were a total of 2776947377 base pair

positions sequenced in this WGS, with 3616955 base pairs that were different to the reference genome (0.13%); of these 38.69% were homozygous changes and 10.61% novel changes (1.56% novel homozygous changes). However novel homozygous changes (56297) represent 0.00202% of bases sequenced in the whole genome. There were a total of 3414278 same-length substitution changes and 40% of these were homozygous. There were a total of 202677 insertion-deletion (indel) changes present (0.007%), of which 57.47% were homozygous. The number of mutations/polymorphisms that were actually located within a gene was 1326149, which represents 0.04% of the WGS. The types of mutations found in gene regions consisted of 8017 synonymous mutations (40.66% homozygous), 7174 missense mutations (40.66% homozygous), 49 nonsense mutations (20.4% homozygous), 5 read-through mutations (40% homozygous), 97 frameshift mutations (77.32% homozygous) and 3 splice site mutations (66.67% homozygous).

Table 7.2.2.1: The non-reference allele (polymorphism/mutations) results

Type of mutation	*Number*	*Percentage*
Total sites mismatching reference	3616955	0.13%
Homozygous total	**1399611**	**38.69%**
Novel total	383887	10.61%
Novel homozygous	**56297**	**1.56%**
Same-length substitution total	3414278	0.12%
Same-length substitution homozygous	**1366503**	**40.00%**
Indel total	202677	0.007%
Indel homozygous	**116489**	**57.47%**
Gene regions mismatching reference (includes novel)	1326149	0.04%
Synonymous total	8017	
Synonymous homozygous	**3260**	**40.66%**
Missense total	7174	
Missense homozygous	**2917**	**40.66%**
Nonsense total	49	
Nonsense homozygous	**10**	**20.4%**
Read-through total	5	
Read-through homozygous	**2**	**40%**
Frameshift	97	
Frameshift homozygous	**75**	**77.32%**
Splice region	3	
Splice region homozygous	**2**	**66.67%**

7.2.2.1. Mutations/Polymorphisms associated with known phenotypes

There were 372938 non-reference alleles (polymorphisms/mutations) in genes that have been reported to be associated with a disease. A total of 5110 of these were in genes that had previously been reported in the literature to be implicated in predisposing to the development of CLL. These 5110 SNPs were in 36 genes (listed in Table 7.2.2.1.1). Table 7.2.2.1.2 lists the genotypes of LK16-1 to specific SNPs known to be associated with an increased predisposition to CLL.

Table 7.2.2.1.1: Summary of genes known to be associated with CLL where non-reference genotypes were identified in LK16-1.

Gene	Location
AKR7A3	Chr1
APC	Chr5
APOC4	Chr19
ATM	Chr11
BRCA2	Chr13
BUB1B	Chr15
CAPN13	Chr2
CFTR	Chr7
CHEK2	Chr22
CLCA2	Chr1
CRY2	Chr11
DKK3	Chr11
DPYD	Chr1

EXOC2	Chr6
FBL1M1	Chr1
FGF2	Chr4
GGT5	Chr22
IL1ORA	Chr11
LPL	Chr8
MCM3AP	Chr21
MIR1302	Chr22
MTR	Chr1
MYO18B	Chr22
NUDT6	Chr4
PDLIM5	Chr4
POLB	Chr8
PPIG	Chr2
RBICC1	Chr8
ROS1	Chr6
SEC14L4	Chr22
SHROOM3	Chr4
TLR4	Chr9
TOP1M1	Chr8
ZC3H3	Chr8
ZNF233	Chr19
ZNF24	Chr18

Table 7.2.2.1.2: Previously identified SNPs known to be associated with an increased risk of CLL where LK16-1 carries the risk allele.

SNP	Gene	Substitution*	Genotype of LK16-1	OR*
Rs10927851	*FBLP-1*	F191S	TT	1.18
Rs997558	*MCM3AP*	S102L	AG	0.86
Rs1048201	*NUDT6*	R209Q	CT	1.20
Rs2250889	*MMP9*	R574P	CC	1.41
Rs1801265	*DPYD*	C29R	AG	0.84
Rs17409304	*CLCA2*	Q306E	CG	0.87
Rs3206824	*DKK3*	G335R	CC	1.16
Rs619203	*ROS1*	S2229C	CG	0.87
Rs459552	*APC*	V1822D	AA	1.15
Rs508405	*CAPN13*	T280A	TT	1.13
Rs1738023	*AKR7A3*	D215N	CC	1.16
Rs1805087	*MTR*	D919G	AG	1.15
Rs2278106	*EPHA7*	P278S	CT	1.34
Rs1211554	*HUS1B*	Y268D	AA	0.79
Rs2032729	*ZNF24*	S220N	CC	1.21
Rs17738527	*SEC14L4*	E211K	CT	0.67
Rs17337252	*RB1CC1*	M234T	AG	1.19
Rs8207	*PPIG*	N699D	AG	0.65

*From Rudd *et al*. 2006(150)

7.2.2.2. Predicted functional mutations identified in LK16-1

There were 154 predicted functional mutations that were detected by the WGS. These are listed in table 7.2.2.2.1. There were a total of 49 nonsense, 97 frameshift, 5 read-through and 3 splice site mutations. 57.79% were homozygous mutations. Four mutations were identified on chromosome 15; two of which were homozygous, only one was located within the region of homozygosity. This was a splice site variation in C15orf48.

Table 7.2.2.2.1: Summary of genetic mutations in protein coding regions

Chr.	Start Position	Ref Allele	LK16-1 Allele 1	LK16-1 Allele 2	Gene(s) mutation
1	12776676	T	A	T	*PRAMEF* nonsense
1	35352095	T	A	T	*ZMYM1* nonsense
1	46853265	G	A	G	*MOBKL2C* nonsense
1	47053333	AT	-	AT	*CYP4B1* Frameshift
1	120138307	C	T	T	*REG4* nonsense
1	143787039	C	C	T	*PDE4DIP* nonsense
1	150462352	T	-	-	*HRNR* Frameshift
1	154831674	-	AC	AC	*GPATCH4* Frameshift
1	156816115	C	T	T	*OR10X1* nonsense
1	159593234	G	A	G	*LOC642502* nonsense; *SDHC* nonsense
1	159742827	C	C	T	*FCGR2A* nonsense

1	169823507	-	C	C	*BAT2L2* Frameshift
1	246179648	T	A	A	*OR2L8* nonsense *OR2L13* nonsense
2	24240683	-	C	C	*LOC375190* Frameshift
2	70916342	-	C	C	*CD207* Frameshift
2	112330899	G	A	G	*ANAPC1* nonsense
2	113710768	-	C	C	*LOC654433* Frameshift *PAX8* Frameshift
2	171889016	A	G	G	*METTL8* read-through
2	179170548	-	G	G	*TTN* Frameshift
2	228184383	C	C	T	*C2orf83* nonsense
3	14172876	G	-	-	*XPC* Frameshift
3	99556281	A	-	-	*OR5K4* Frameshift
3	99593097	-	A	A	*OR5K3* Frameshift
3	155440980	CAGT	-	-	*SGEF* Frameshift
3	161501501	-	A	-	*IFT80* Frameshift
4	75996	-	A	A	*ZNF718* Frameshift *ZNF595* Frameshift
4	76102	-	C	C	*ZNF718* Frameshift *ZNF595* Frameshift

4	77314	-	A	A	*ZNF718* Frameshift *ZNF595* Frameshift
4	22429601	A	A	T	*GBA3* read-through
4	47633340	G	A	G	*CNGA1* nonsense
4	70547375	A	A	T	*UGT2A1* nonsense
4	104052059	G	A	G	*NHEDC1* nonsense
4	129061252	A	A	C	*MFSD8* nonsense
4	152420469	-	CAGGT	CAGGT	*PRSS48* Frameshift
4	155463851	TTTG	-	-	*DCHS2* Frameshift
4	185787625	TTGAA	-	TTGAA	*CASP3* Frameshift
5	1293756	C	G	G	*SLC6A18* nonsense
5	61911470	-	G	G	*LRRC70* Frameshift *IPO11* Frameshift
5	101625533	G	G	T	*SLCO4C1* nonsense
5	135540985	-	C	C	*SMAD5* Frameshift
5	140548219	-	A	A	*PCDHB9* Frameshift
5	149355072	T	-	-	*TIGD6* Frameshift
5	156654442	-	C	C	*CYFIP2* Frameshift
6	56589492	AT	-	AT	*DST* Frameshift
6	65074936	GA	-	-	*EYS* Splice site variation
6	117256700	G	A	G	*GPRC6A* nonsense
6	167629691	G	A	G	*UNC93A* nonsense

7	12357794	-	A	-	*VWDE* Frameshift
7	47921711	-	A	A	*PKD1L1* Frameshift
7	93378079	C	C	T	*GNGT1* nonsense
8	3254411	C	-	-	*CSMD1* Frameshift
8	6660787	A	-	-	*XKR5* Frameshift
8	24866981	G	-	-	*NEFL* Frameshift
8	30740383	-	T	T	*UBXN8* Frameshift
8	37913463	-	T	T	*GOT1L1* Frameshift
8	38946342	C	-	-	*PLEKHA2* Frameshift
8	48968370	-	G	G	*PRKDC* Frameshift
8	64261284	-	G	G	*YTHDF3* Frameshift
8	126088870	G	-	-	*SQLE* Frameshift
9	106400589	T	-	-	*OR13C5* Frameshift
9	106407485	GC	-	-	*OR13C2* Frameshift
9	124431592	-	A	A	*OR1B1* Frameshift
9	131631329	A	A	G	*C9orf78* read-through
9	138754315	G	A	G	*LCN10* nonsense
10	7645083	C	-	-	*ITIH5* Frameshift
10	27727230	A	G	G	*PTCHD3* read-through
10	50360764	C	A	C	*ERCC6* nonsense
10	97910090	-	C	C	*ZNF518A* Frameshift
10	118373454	-	G	G	*PNLIPRP2* Frameshift
10	126717591	T	A	T	*CTBP2* nonsense

11	4345980	G	-	-	*OR52B4* Frameshift
11	4747449	G	-	-	*OR51F1* Frameshift
11	5129372	-	C	C	*OR52A1* Frameshift
11	5733059	A	T	T	*OR52N4* nonsense
11	14058070	-	C	C	*SPON1* Frameshift
11	46298836	-	G	-	*CREB3L1* Frameshift
11	48242806	T	A	A	*OR4X1* nonsense
11	48303589	G	A	G	*OR4C3* nonsense
11	55096227	C	C	T	*OR4C16* nonsense
11	56187791	C	T	T	*OR5AR1* nonsense *OR8U8* nonsense
11	59921928	TT	-	-	*MS4A14* Frameshift
11	60021577	C	T	T	*MS4A12* nonsense
11	60922308	-	A	A	*TMEM216* Frameshift
11	62814500	G	A	G	*SLC22A10* nonsense
11	67542641	-	C	C	*ALDH3B1* Frameshift
11	94339156	-	T	T	*CWC15* Frameshift
11	94440551	-	C	C	*SFRS2B* Frameshift
11	99195585	C	C	T	*CNTN5* nonsense
11	111358317	-	C	C	*DIXDC1* Frameshift
11	118403645	C	-	-	*SLC37A4* Frameshift
11	118445150	-	C	C	*VPS11* Frameshift
11	124957511	-	C	C	*EI24* Frameshift

12	892829	A	A	C	*RAD52* nonsense
12	10162353	A	A	C	*CLEC7A* nonsense
12	46883143	-	A	-	*OR10AD1* Frameshift
12	53809853	T	-	T	*OR9K2* Frameshift
12	53927521	C	C	T	*OR6C74* nonsense
12	54107225	A	-	-	*OR6C76* Frameshift
12	84162776	A	-	-	*LRRIQ1* Frameshift
12	100943940	-	G	G	*CCDC53* Frameshift
12	111376815	T	G	T	*PTPN11* nonsense
14	19736016	-	A	-	*OR11G2* Frameshift
14	62854161	-	G	G	*GPHB5* Frameshift
15	39792871	C	C	T	*MGA* nonsense
15	43510320	AAGT	-	-	*C15orf48* Splice site variation
15	63704669	C	-	C	*SLC24A1* Frameshift
15	81145551	-	C	C	*AP3B2* Frameshift
16	3542228	G	-	-	*NLRC3* Frameshift
16	31678198	-	A	-	*ZNF720* Frameshift
16	57134816	A	-	A	*CNOT1* Frameshift
16	70538916	-	TTTG	-	*PKD1L3* Frameshift
16	74869104	-	T	T	*CNTNAP4* Frameshift
16	79799698	G	A	G	*PKD1L2* nonsense
16	80591310	G	A	G	*SDR42E1* nonsense

17	3541025	G	-	G	*P2RX5* Frameshift
17	6496271	-	G	-	*C17orf100* Frameshift
17	7411009	A	-	-	*SENP3* Frameshift
17	7435886	C	-	-	*FXR2* Frameshift
17	8665940	-	G	G	*PIK3R6* Frameshift
17	21144802	C	C	T	*MAP2K3* nonsense
17	23716352	-	G	-	*SEBOX* Frameshift
17	23751848	A	-	-	*SLC46A1* Frameshift *SARM1* Frameshift
17	24110580	-	G	-	*C17orf63* Frameshift
17	31124464	-	C	-	*MMP28* Frameshift
17	40720077	-	G	-	*MAP3K14* Frameshift
17	42602360	G	A	G	*CDC27* nonsense
17	57022734	G	A	G	*NACA2* nonsense
17	59014621	G	-	-	*DCAF7* Frameshift
17	71589391	C	C	T	*ZACN* nonsense *EXOC7* nonsense
17	77225337	TAAC	-	*TAAC*	*TSPAN10* Splice site variation
18	31980332	A	-	A	*ELP2* Frameshift
18	50134886	G	A	G	*STARD6* nonsense
19	15591499	-	C	-	*CYP4F8* Frameshift
19	40410859	C	T	T	*FAM187B* nonsense

19	40906473	-	G	G	*MLL4* Frameshift
19	45814934	-	G	G	*LTBP4* Frameshift
19	49708830	T	-	-	*CEACAM20* Frameshift
19	56420915	CCGG	-	-	*CD33* Frameshift
19	56527705	-	G	G	*VSIG10L* Frameshift
19	56696714	G	A	G	*SIGLEC12* nonsense
19	57495481	TG	-	-	*ZNF480* Frameshift
19	61191090	G	C	G	*NLRP8* read-through
19	62334593	C	A	A	*USP29* nonsense
20	74310	CC	-	-	*DEFB126* Frameshift
20	21134162	-	G	G	*PLK1S1* Frameshift
21	30835852	G	-	-	*KRTAP19* Frameshift
21	31123840	A	-	-	*KRTAP7* Frameshift
21	33088058	-	T	T	*C21orf62* Frameshift *C21orf49* Frameshift
21	33088060	A	A	T	*C21orf62* nonsense *C21orf49* nonsense
22	34917792	-	CT	CT	*APOL4* Frameshift
22	44188943	-	C	C	*RIBC2* Frameshift
X	50137887	-	G		*DGKK* Frameshift
X	69395896	C	T		*P2RY4* nonsense
X	151659500	C	G		*CSAG1* nonsense

7.2.2.3. Mutations identified within the regions of IBD identified in Chapter 5

The top 3 regions where the most number of pairs share genotypes were on chromosome 3 (Figure 7.2.2.3.1), chromosome 7 (Figure 7.2.2.3.2) and chromosome 9 (Figure 7.2.2.3.3). These regions were identified by the PLINK analysis as potentially harbouring predisposition alleles in the 13 Tasmanian families.

Figure 7.2.2.3.1: Number of pairs sharing genotypes on chromosome 3

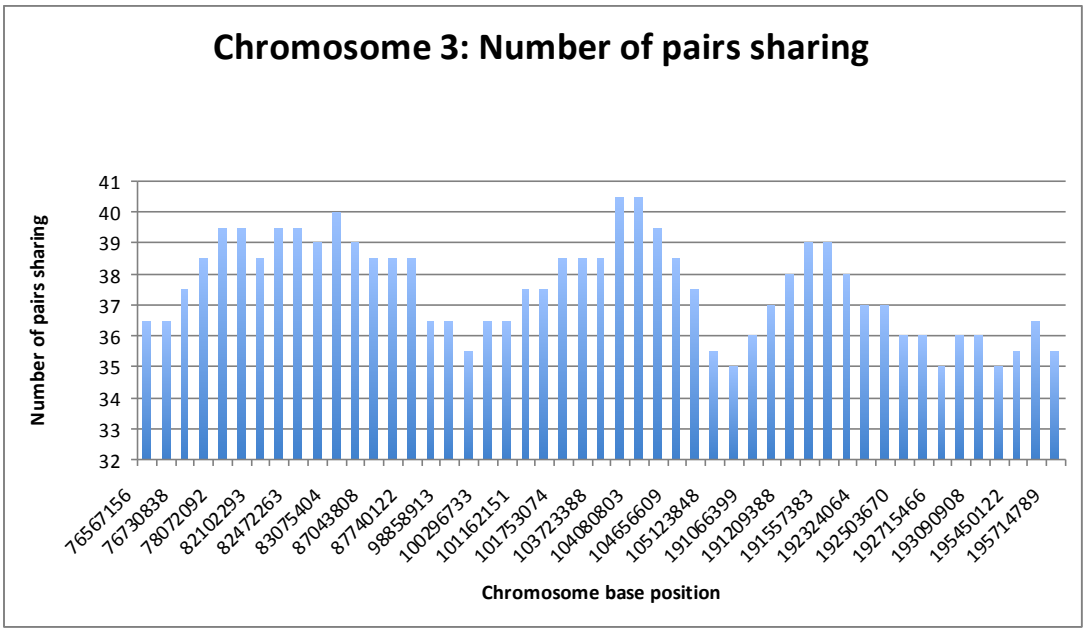


Figure 7.2.2.3.2: Number of pairs sharing genotypes on chromosome 7

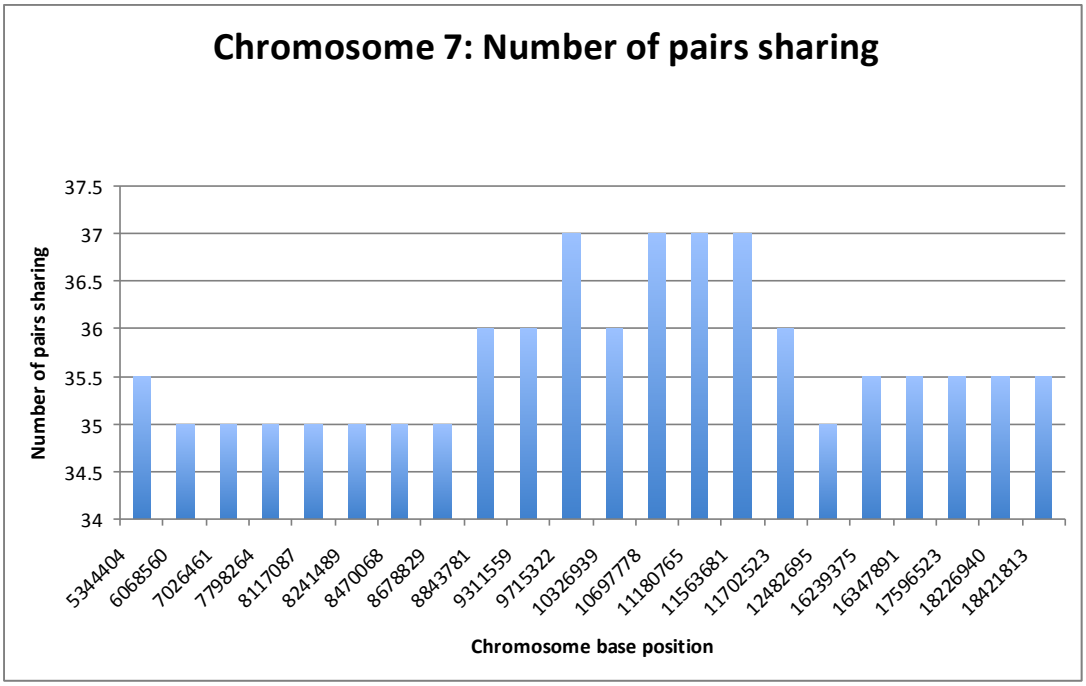
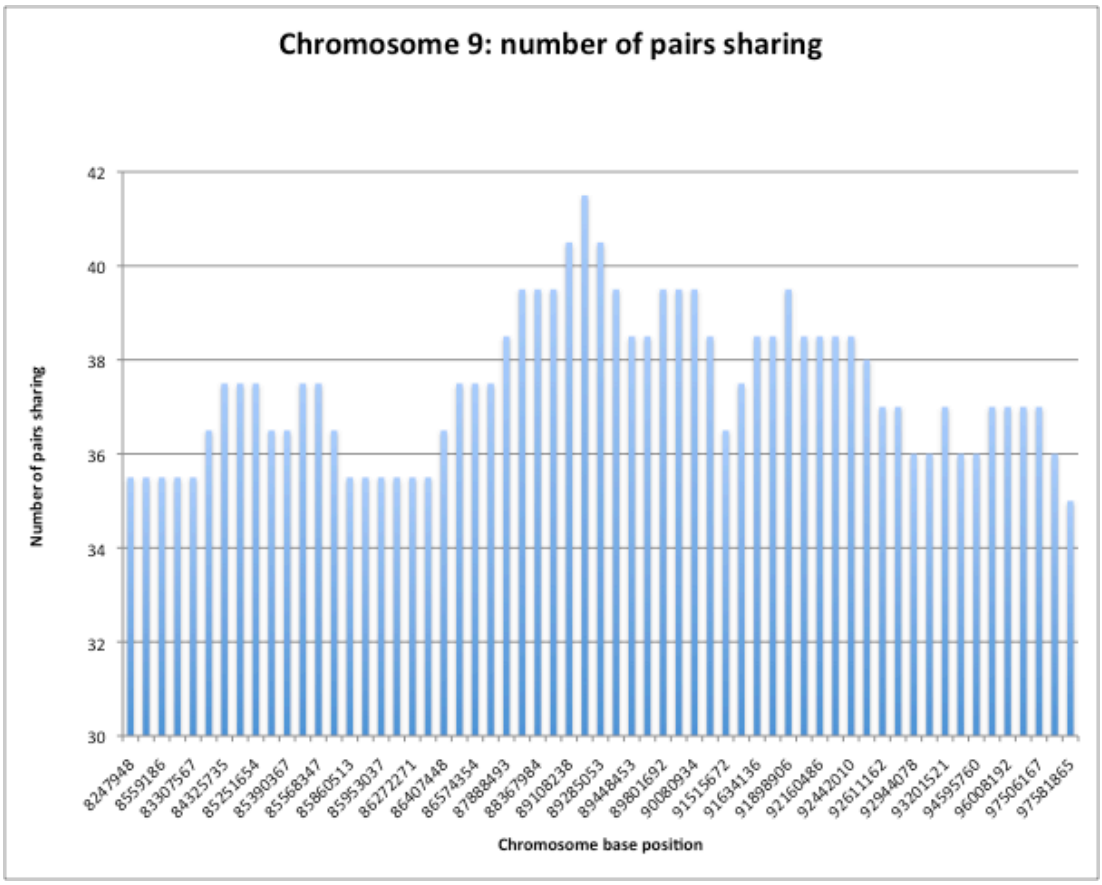


Figure 7.2.2.3.3: Number of pairs sharing genotypes on chromosome 9



There are two predicted functional mutations within the region on chromosome 3 and one mutation within the region on chromosome 7 (Table 7.2.2.3.1). No predicted functional mutations were identified within the chromosome 9 region. Confirming that no mutation was present anywhere near or in the coding regions of *DAPK1*.

Table 7.2.2.3.1: Predicted functional mutations in these regions in LK16-1

Chr.	Start Position	Ref	Allele 1	Allele 2	Gene
3	99556281	A	-	-	*OR5K4* Frameshift
3	99593097	-	A	A	*OR5K3* Frameshift
7	12357794	-	A	-	*VWDE* Frameshift

7.2.2.4. Novel mutations identified

There were 33 novel predicted functional mutations identified in this WGS of LK16-1, of which 10 were homozygous mutations (Table 7.2.2.4.1).

Table 7.2.2.4.1: List of the novel predicted functional mutations identified in LK16-1

Chr.	start	end	ref	allele	allele	Gene(s)
1	35352095	35352096	T	A	T	*ZMYM1*
1	150462352	150462353	T	-	-	*HRNR†*
1	159593234	159593235	G	A	G	*LOC642502/SDHC*
3	99593097	99593097	-	A	A	*OR5K3**
3	161501501	161501501	-	A	-	*IFT80*
4	75996	75996	-	A	A	*ZNF718/ZNF595**
4	47633340	47633341	G	A	G	*CNGA1*
4	70547375	70547376	A	A	T	*UGT2A1*
4	104052059	104052060	G	A	G	*NHEDC1*
4	129061252	129061253	A	A	C	*MFSD8*
4	155463851	155463855	TTTG	-	-	*DCHS2**
4	185787625	185787630	TTGAA	-	TTGAA	*CASP3*
5	101625533	101625534	G	G	T	*SLCO4C1*
6	56589492	56589494	AT	-	AT	*DST*
7	47921711	47921711	-	A	A	*PKD1L1**
7	93378079	93378080	C	C	T	*GNGT1*
9	124431592	124431592	-	A	A	*OR1B1**
10	50360764	50360765	C	A	C	*ERCC6*
10	126717591	126717592	T	A	T	*CTBP2†*
11	4345980	4345981	G	-	-	*OR52B4**

11	5129372	5129372	-	C	C	*OR52A1**
11	124957511	124957511	-	C	C	*EI24**
12	53809853	53809854	T	-	T	*OR9K2†*
12	111376815	111376816	T	G	T	*PTPN11†*
15	39792871	39792872	C	C	T	*MGA*
15	43510320	43510324	AAGT	-	-	*C15ORF48**
16	3542228	3542229	G	-	-	*NLRC3* †*
16	31678198	31678198	-	A	-	*ZNF720*
17	6496271	6496271	-	G	-	*C17ORF100*
19	15591499	15591499	-	C	-	*CYP4F8†*
19	56420915	56420919	CCGG	-	-	*CD33*
21	33088060	33088061	A	A	T	*C21ORF62/C21ORF49*
X	69395896	69395897	C	T		*P2RY4*

* novel homozygous mutations

† mutation has been reported in the 1000 genomes project

7.3. Discussion and Conclusion

One of the limitations of familial studies has been that the ability to find the causative genetic mutations depends upon having a minimum of 10 affected individuals to identify a LOD score (is the decimal log likelihood ratio) greater than 3. Now with next generation sequencing this requirement is no longer necessary and WGS has been used in Miller syndrome to identify the causative gene(235). WGS has also been used in 10 unrelated Kabuki syndrome patients to explain the variation

in the clinical symptoms of these patients and a common mutation in *MLL2* was discovered(236).

Ley *et al*. 2008 identified 14 variants in their WGS of cytogenetically normal AML cells, 4 were inherited variants and 10 were acquired variants in coding regions(160). Two acquired mutations were in genes commonly mutated in AML (*FLT3* and *NPM1*), the other genes were *PTPRT*, *CDH24*, *PCLKC*, *SLC15A1*, *KNDC1*, *GPR123*, *EBI2* and *GRINL1B*. Four of these genes (*PTPRT*, *CDH24*, *PCLKC* and *SLC15A1*) belong to gene families that are strongly associated with cancer pathogenesis and the remaining four have not been previously implicated in leukaemia(160). The WGS has been reanalysed and another mutation was identified in *DNMT3A,* this has subsequently been identified as a commonly mutated gene in cytogenetically normal AML and is associated with a poor prognosis(237). They also reported that all the acquired mutations, except the *FLT3* mutations were heterozygous and present in virtually all the tumour cells(160).

As the DNA that was used in the sequencing of LK16-1 represented tumour cells, therefore some of the mutations identified would be acquired mutations in his CLL cells, rather than germline mutations. However, due to the family structure of the LK16 family with his parents being first cousins, a homozygous mutation is implicated in this family. There were a total of 89 predicted homozygous predicted functional mutations identified. All mutations present in LK16-1 will be assessed in the remaining family members to confirm whether they are a false positive, germline or an acquired mutation. Knowledge of the genome-wide SNP array results in the LK16 family, especially from LK16-106 (child of LK16-1) and LK16-4 (brother of LK16-1), allows assessment of whether the mutations are germline (if homozygous

in LK16-1, LK16-106 is an obligate carrier; if heterozygous there is a 50% chance that LK16-4 or LK16-106 is a carrier; and it is known which regions of the genome these people share with LK16-1). The germline mutations should be present in the other family members, where as acquired mutations in LK16-1's CLL cells should not be.

The confirmed acquired mutations will also be assessed in other patients who have CLL, both in their CLL cells (peripheral blood DNA) and their germline DNA (extracted from saliva samples). These acquired mutations have the potential to provide knowledge on the mutations required for the development of CLL and therefore may provide targets for novel therapies.

The WGS identified 372938 non-reference alleles present in gene regions in LK16-1. 5110 were in 36 genes that have been reported in the literature to predispose to the development of CLL(150). Rudd *et al.* 2006, reported 49 germline SNPs in the *ATM-BRCA2-CHEK2* axis that are significantly associated with the risk of developing CLL(150). It was found that LK16-1 carries the risk allele for 18 of these SNPs, of which 8 are in the homozygous state.

Of the genes where a predicted functional mutation was identified, 20 are involved in DNA repair (*PRKDC*, *ERCC6* and *RAD52*), cell cycle (*MOBKL2C*, *PDE4DIP*, *ANAPC1*, *MGA*, *CDC27* and *PLKIS1*), tumourogenesis (*MMP28*, *DIXDC1*, *CSMD1*, *PAX8*, *MLL4*, *ITIH5* and *C15orf48*) or apoptosis pathways (*CASP3*, *CYFIP2*, *EI24* and *CD33*). Ten of these were heterozygous mutations and 11 were homozygous mutations. Intriguingly the homozygous mutations were mostly in genes involved in the apoptosis or tumourogenesis pathways, suggesting that genes in these pathways act as tumour suppresser genes.

The WGS has identified 154 predicted functional mutations in LK16-1's CLL cells, 89 of these were homozygous mutations. Only one homozygous mutation was identified within the larger region of interest on chromosome 15, this was a splice site variation in *C15orf48*. Very little information is known about *C15orf48*, but it was first identified in a study of oesophageal carcinomas(238). In this study it was proposed that *C15orf48* was a tumour suppressor gene(238). This would fit with the hypothesis in the LK16 family.

The regions of the genome identified in chapter 5 that showed the most sharing in the IBD analysis were on chromosome 9 (8247948-97581865), chromosome 3 (188583591-199322659; 73482376-105281251; 39653716-58867074) and chromosome 7 (9945064-10945597). Two homozygous frameshift mutations in two genes (*OR5K3* and *OR5K4*) were identified in the chromosome 3 region, both belonging to the olfactory receptors family of genes. A heterozygous frameshift mutation was present in the *VWDE* gene located within the chromosome 7 region. No predicted functional mutations were identified in genes within the chromosome 9 region. None of these genes (*VWDE*, *OR5K3* or *ORK4*) have been implicated previously as having a role in leukaemogenesis.

According to the 1000 genomes browser (http://browser.1000genomes.org/Homo_sapiens/Gene/Variation_Gene/Table?db=core;g=ENSG00000166920;r=15:43510055-43512937), which claims to provide over 95% of the novel variants in the human genome(234), the splice site variant identified in *C15orf48* has not been found in their dataset. This supports the possibility that this is a very rare mutation and may be the causative mutation in this family. The 1000 genome project has in fact identified six of the novel mutations that

were identified in LK16-1. Many of the novel mutations identified in LK16-1 may be due to the fact that some of the novel mutations in LK16-1 are acquired mutations in his CLL cells. However, these mutations were also assessed in the publically available data from the cancer genome project (http://www.sanger.ac.uk/genetics/CGP/cosmic/) and they were also not present there either, although this project does not include CLL as one of the specific cancers that they are sequencing, they are sequencing AML genomes.

WGS of LK16-1 has potentially identified a novel mutation in *C15orf48*. This, as well as the remaining predicted functional mutations is in the process of being confirmed in the remaining members of LK16 family, the other 12 Tasmanian families and a sporadic dataset of 100 patients with haematological malignancies.

8. Chapter 8: Summary and Conclusions

8.1. Current aims

The aim of this study was to identify the genetic mutation(s) in these 13 Tasmanian families. These families have been expanded to include the current generations and all those affected have had their disease verified and reclassified into the current haematological classification system(1). The hypothesis is that there is a predisposition allele in these families; and due to the structure of some of the families some of the affected have inherited two copies of the predisposition allele. More families are still being referred to the study and are being consented to participate. These families are being systematically cross referenced with individuals on the TCR to identify new cases in these families.

8.2. Familial Haematological Malignancy pedigree collection.

Thirteen families were selected for the study, due to their high number of affected individuals. A total of 130 affected individuals were identified in these 13 Tasmanian families. Nine families show a statistically significant reduction in the age at diagnosis for subsequent generations (anticipation), and two show a pattern of inheritance that would be in keeping with a recessive mutation(162). Another family (LK124) also shows a cluster of a rare subtype of DLBCL, with three people having primary CNS lymphoma. These Tasmanian families have a denser aggregation of affected members than families in the literature. DNA from three members of the LK124 family from different generations is being used in a whole genome sequence to assess the size of repeats (dynamic mutations) across generations.

8.3. Investigation of previously identified predisposing mutations

Only one causative mutation has been found in one family with familial CLL in the literature(25). This was a mutation upstream of *DAPK1*. This mutation, 3'UTR, 5'UTR, cDNA, splices site variants and 500bp upstream and downstream of *DAPK1* were sequenced in these 13 Tasmanian families and found to be normal.

8.4. High density SNP array of LK16 CLL family.

This identified a region on chromosome 15 where the 5 affected siblings are homozygous.

8.5. High density SNP array of all 13 Tasmanian families

This used DNA from 117 people in these 13 families, representing 78 affected individuals. CNV analysis revealed how polymorphic the human genome is. One CNV on chromosome 4q13 was shared in the samples from tumour DNA (CLL samples), were they all had 3 copies of this region. This CNV has a marked difference is allele frequency in different ethnic populations correlating with the incidence of CLL. PLINK IBD analysis of the regions of sharing by individuals showed the regions with the most sharing are on chromosome 7, chromosome 3 and chromosome 9. It was interesting that *DAPK1* is located in the region on chromosome 9, however no mutation was found in this gene. Two predicted functional novel mutations were identified in genes in the chromosome 3 region and one in the chromosome 7 region. SNP analysis looking for increased association of

SNPs in the affected compared too non-affected controls highlighted several SNPs of significance. Formal linkage analysis of these 13 families in currently also being completed.

8.6. Cell lines and gene expression

This identified *RNF111* as the most differentially expressed gene and sequencing identified a missense mutation in exon 2 of this gene. The genotype of this missense mutation that the LK16 family carry is common in western countries, but rare in Asia, correlating with the incidence of CLL in these two geographical regions. However this missense mutation is considered a normal variant, due to its high allele frequency. This missense mutation is being followed up in Asian CLL patients.

8.7. Whole genome sequencing of LK16-1

Whole genome sequencing was undertaken on LK16-1. This was done on an Illumina platform with greater than 30 times coverage of the genome. A total of 154 predicted functional mutations were identified. Only one of these was within the larger region of homozygosity on chromosome 15. This was a homozygous splice site variation (4 base pair deletion) in C15orf48. This splice site mutation has not been reported in any other databases of human variation. This gene was first identified in oesophageal carcinoma as a potential tumour suppressor gene(238). These 154 predicted functional mutations are being confirmed in the all the familial dataset (250 people) and 100 sporadic HM controls to assess whether they are a false positive (error), acquired mutation or a germline mutation.

Importantly this is the first WGS of a CLL genome. Many of the mutations identified also represent acquired changes in the CLL cells and provide insight into the genetic changes that occur for this disease to develop. These may potentially provide new targets for therapy.

8.8. Conclusion

This study has identified and documented 13 Tasmanian families with a dense aggregation of people with familial HMs. A putative predisposition haplotype was identified in the LK16 family on chromosome 15 and the potential predisposition mutation in this family identified in *C15orf48*. The WGS also identified many acquired mutations that occur in the development of CLL. These are currently being confirmed in the familial and sporadic dataset of HMs. This study has identified several novel and known alleles for the genetic basis of HMs and is consistent with the literature in that this is a complex genetic disorder and many common genetic polymorphisms are responsible for the heritable predisposition in HMs.

9. References:

1. WHO classification of Tumours of Haematopoietic and Lymphoid Tissues. 4th ed. Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, et al., editors: International Agency for Research on Cancer (IARC), Lyon France; 2008.

2. Dalton M, Venn A, Albion T, Otahal P, Blizzard L. Cancer in Tasmania: Incidence and Mortality 2005. Menzies Research Institute. 2008.

3. Chang ET, Smedby KE, Hjalgrim H, Porwit-MacDonald A, Roos G, Glimelius B, et al. Family History of Hematopoietic Malignancy and Risk of Lymphoma. The Journal of the National Cancer Institute. 2005;97(19):1466-74.

4. Crowther-Swanepoel D, Houlston RS. The molecular basis of familial chronic lymphocytic leukemia. Haematologica. 2009;94(5):606-9.

5. Deshpande HA, Hu XP, Marino P, Jan NA, Wiernik PH. Anticipation in familial plasma cell dyscrasias. British Journal of Haematology. 1998;103(3):696-703.

6. Goldin LR, Pfeiffer RM, Li X, Hemminki K. Familial risk of lymphoproliferative tumors in families of patients with chronic lymphocytic leukemia: results from the Swedish Family-Cancer Database. Blood. 2004;104(6):1850-4.

7. Goldin LR, Sgambati M, Marti GE, Fontaine L, Ishibe N, Caporaso N. Anticipation in Familial Chronic Lymphocytic Leukemia. The American Journal of Human Genetics. 1999;65(1):265-8.

8. Goldin LR, Slager SL. Familial CLL: Genes and Environment. Hematology. 2007(1):339-45.

9. Grosbois B, Jego P, Attal M, Payen C, Rapp MJ, Fuzibet JG, et al. Familial multiple myeloma: report of fifteen families. British Journal of Haematology. 1999;105(3):768-70.

10. Horwitz M, Goode EL, Jarvik GP. Anticipation in Familial Leukemia. American Journal of Human Genetics. 1996;59:990-8.

11. Houlston RS, Sellick G, Yuille M, Matutes E, Catovsky D. Causation of chronic lymphocytic leukemia - Insights from familial disease. Leukemia Research. 2003;27(10):871-6.

12. Kurita S, Kamei Y, Ota K. Genetic studies on familial leukaemia. Cancer. 1974;34(4):1098-101.

13. Lynch HT, Ferrara K, Barlogie B, Coleman EA, Lynch JF, Weisenburger D, et al. Familial Myeloma. New England Journal of Medicine. 2008;359(2):152-7.

14. Lynch HT, Sanger WG, Pirruccello S, Quinn-Laquer B, Weisenburger DD. Familial Multiple Myeloma: a Family Study and Review of the Literature. The Journal of the National Cancer Institute. 2001;93(19):1479-83.

15. Mauro FR, Giammartini E, Gentile M, Sperduti I, Valle V, Pizzuti A, et al. Clinical features and outcome of familial chronic lymphocytic leukemia. Haematologica. 2006;91(8):1117-20.

16. Owen C, Barnett M, Fitzgibbon J. Familial myelodysplasia and acute myeloid leukaemia - a review. British Journal of Haematology. 2008;140(2):123-32.

17. Segel GB, Lichtman MA. Familial (inherited) leukemia, lymphoma, and myeloma: an overview. Blood Cells, Molecules, and Diseases. 2004;32(1):246-61.

18. Sellick GS, Goldin LR, Wild RW, Slager SL, Ressenti L, Strom SS, et al. A high-density SNP genome-wide linkage search of 206 families identifies susceptibility loci for chronic lymphocytic leukemia. Blood. 2007;110(9):3326-33.

19. Shpilberg O, Modan M, Modan B, Chetrit A, Fuchs Z, Ramot B. Familial aggregation of haematological neoplasms: a controlled study. British Journal of Haematology. 1994;87(1):75-80.

20. Shugart YY, Hemminki K, Vaittinen P, Kingman A, Dong C. A genetic study of Hodgkin's lymphoma: an estimate of heritability and anticipation based on the familial cancer database in Sweden. Human Genetics. 2000;106(5):553-6.

21. Smith ML, Cavenagh JD, Lister TA, Fitzgibbon J. Mutation of CEBPA in Familial Acute Myeloid Leukemia. New England Journal of Medicine. 2004;351(23):2403-7.

22. Wiernik P.H., Ashwin M., Xiao-Ping HXP, Paietta E., Brown K. Anticipation in familial chronic lymphocytic leukaemia. British Journal of Haematology. 2001;113(2):407-14.

23. Wiernik PH, Wang SQ, Hu X, Marino P, Paietta E. Age of onset evidence for anticipation in familial non-Hodgkin's lymphoma. British Journal of Haematology. 2000;108(1):72-9.

24. Yuille MR, Matutes E, Marossy A, Hilditch B, Catovsky D, Houlston RS. Familial chronic lymphocytic leukaemia: a survey and review of published studies. British Journal of Haematology. 2000(109):794-9.

25. Raval A, Tanner SM, Byrd JC, Angerman EB, Perko JD, Chen S-S, et al. Downregulation of Death-Associated Protein Kinase 1 (DAPK1) in Chronic Lymphocytic Leukemia. Cell. 2007;129(5):879-90.

26. Donald M. Concise Review: Hematopoietic Stem Cells and Tissue Stem Cells: Current Concepts and Unanswered Questions. Stem Cells. 2007;25(10):2390-5.

27. Wen D, Wang S, Zhang L, Wei L, Zhou W, Peng Q. Early onset, multiple primary malignancies, and poor prognosis are indicative of an inherited predisposition to esophageal squamous cell carcinoma for the familial as opposed to the sporadic cases - An update on over 14-year survival. European Journal of Medical Genetics. 2009;52(6):381-5.

28. Alfred G. Knudson J. Mutation and Cancer: Statistical Study of Retinoblastoma Proceedings of the National Academy of Sciences. 1971;68(4):820-3.

29. Domenico M, Sonia De F, Aldo Di L, Laura L, Theodora H. Does the evidence matter in medicine? The retinoblastoma paradigm. International Journal of Cancer. 2007;121(11):2501-5.

30. Erenpreisa J, Cragg MS. Cancer: A matter of life cycle? Cell Biology International. 2007;31(12):1507-10.

31. Yaddanapudi R. Down syndrome and leukemia: New insights into the epidemiology, pathogenesis, and treatment. Pediatric Blood & Cancer. 2005;44(1):1-7.

32. Goldin LR, Bjorkholm M, Kristinsson SY, Turesson I, Landgren O. Elevated risk of chronic lymphocytic leukemia and other indolent non-Hodgkin's lymphomas among relatives of patients with chronic lymphocytic leukemia. Haematologica. 2009;94(5):647-53.

33. Giles GG, Lickiss JN, Baikie MJ, Lowenthal RM, Panton J. Myeloproliferative and lymphoproliferative disorders in Tasmania, 1972-80: occupational and familial aspects. The Journal of the National Cancer Institute 1984;72(6):1233-40.

34. Catovsky D. Definition and diagnosis of sporadic and familial chronic lymphocytic leukemia. Hematology/Oncology Clinics of North America. 2004;18(4):783-94.

35. Gale RP, Cozen W, Goodman MT, Wang FF, Bernstein L. Decreased chronic lymphocytic leukemia incidence in Asians in Los Angeles County. Leukemia Research. 2000;24(8):665-9.

36. Belson M, Kingsley B, and Holmes A. Risk Factors for Acute Leukemia in Children: A Review. Environ Health Perspect 2007;115(1):138-45.

37. Matti H. Epidemiologic evidence for multi-stage theory of carcinogenesis. International Journal of Cancer. 1971;7(3):557-64.

38. Morris JA. A mutational theory of leukaemogenesis. Journal of Clinical Pathology. 1989 April 1, 1989;42(4):337-40.

39. Kralovics R. Genetic complexity of myeloproliferative neoplasms. Leukemia. 2008;22(10):1841-8.

40. David R. Stem cells: Imaging the origin of HSCs. Nature Reviews Molecular Cell Biology. 2010;11:232-3.

41. Orkin SH, Zon LI. Hematopoiesis: An Evolving Paradigm for Stem Cell Biology. Cell. 2008;132(4):631-44.

42. World health Organization classification of Tumours; Pathology and Genetics: Tumours of Haematopoietic and lymphoid tissues. Jaffe ES, Harris NL, Stein H, JW V, editors: IARC Press; 2001.

43. Kuendgen A, Lübbert M. Current status of epigenetic treatment in myelodysplastic syndromes. Annals of Hematology. 2008;87(8):601-11.

44. Issa J-P. Epigenetic Changes in the Myelodysplastic Syndrome. Hematology/Oncology Clinics of North America. 2010;24(2):317-30.

45. Dameshek W. Editorial: Some Speculations on the Myeloproliferative Syndromes. Blood. 1951 April 1, 1951;6(4):372-5.

46. Goldman JM, Melo JV. Targeting the BCR-ABL Tyrosine Kinase in Chronic Myeloid Leukemia. New England Journal of Medicine. 2001 April 5, 2001;344(14):1084-6.

47. Nowell P, Hungerford, D. A minute chromosome in human chronic granulocytic leukemia. Science. 1960;132:1497.

48. Rowley JD. A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature. 1973;243(5405):290-3.

49. Druker BJ, Lydon NB. Lessons learned from the development of an Abl tyrosine kinase inhibitor for chronic myelogenous leukemia. The Journal of Clinical Investigation. 2000;105(1):3-7.

50. Kralovics R, Passamonti F, Buser AS, Teo S-S, Tiedt R, Passweg JR, et al. A Gain-of-Function Mutation of JAK2 in Myeloproliferative Disorders. New England Journal of Medicine. 2005;352(17):1779-90.

51. Vannucchi AM, Antonioli E, Guglielmelli P, Pardanani A, Tefferi A. Clinical correlates of JAK2V617F presence or allele burden in myeloproliferative neoplasms: a critical reappraisal. Leukemia. 2008;22(7):1299-307.

52. Pabst T, Eyholzer M, Haefliger S, Schardt J, Mueller BU. Somatic CEBPA Mutations Are a Frequent Second Event in Families With Germline CEBPA Mutations and Familial Acute Myeloid Leukemia. Journal of Clinical Oncology. 2008;26(31):5088-93.

53. Michaud J, Wu F, Osato M, Cottles GM, Yanagida M, Asou N, et al. In vitro analyses of known and novel RUNX1/AML1 mutations in dominant familial platelet disorder with predisposition to acute myelogenous leukemia: implications for mechanisms of pathogenesis. Blood. 2002;99(4):1364-72.

54. Agnelli L, Bicciato S, Mattioli M, Fabris S, Intini D, Verdelli D, et al. Molecular Classification of Multiple Myeloma: A Distinct Transcriptional Profile Characterizes Patients Expressing CCND1 and Negative for 14q32 Translocations. Journal of Clinical Oncology. 2005;23(29):7296-306.

55. Huang JZ, Sanger WG, Greiner TC, Staudt LM, Weisenburger DD, Pickering DL, et al. The t(14;18) defines a unique subset of diffuse large B-cell lymphoma with a germinal center B-cell gene expression profile. Blood. 2002;99(7):2285-90.

56. Campbell LJ. Cytogenetics of lymphomas. Pathology. 2005;37(6):493-507.

57. Dave SS, Fu K, Wright GW, Lam LT, Kluin P, Boerma E-J, et al. Molecular Diagnosis of Burkitt's Lymphoma. New England Journal of Medicine. 2006;354(23):2431-42.

58. Mitelman F, Mertens F, Johansson B. A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. Nat Genet. 1997;15(4s):417-74.

59. Goldman JM, Melo JV. BCR-ABL in Chronic Myelogenous Leukemia , How Does It Work? Acta Haematologica. 2008;119(4):212-7.

60. Faderl S, Talpaz M, Estrov Z, O'Brien S, Kurzrock R, Kantarjian HM. The Biology of Chronic Myeloid Leukemia. New England Journal of Medicine. 1999;341(3):164-72.

61. Thirman MJ, Gill HJ, Burnett RC, Mbangkollo D, McCabe NR, Kobayashi H, et al. Rearrangement of the MLL Gene in Acute Lymphoblastic and Acute Myeloid Leukemias with 11q23 Chromosomal Translocations. New England Journal of Medicine. 1993;329(13):909-14.

62. Rodriguez-Perales S, Cano F, Lobato M, Rabbitts T. MLL gene fusions in human leukaemias: in vivo modelling to recapitulate these primary tumourigenic events. International Journal of Hematology. 2008;87(1):3-9.

63. Jansen MWJC, van der Velden VHJ, van Dongen JJM. Efficient and easy detection of MLL-AF4, MLL-AF9 and MLL-ENL fusion gene transcripts by multiplex real-time quantitative RT-PCR in TaqMan and LightCycler. Leukemia. 2005;19(11):2016-8.

64. Emanuela M, Fiorenza A, Fabrizio V, Franco L, Giovanni A, Adriana Z, et al. Trisomy 8 in myelodysplasia and acute leukemia is constitutional in 15-20% of cases. Genes, Chromosomes and Cancer. 2002;33(1):93-7.

65. Cancer cytogenetics. 3rd ed. Heim S, Mitelman F, editors. Hoboken, New Jersey: John Wiley and Sons; 2009.

66. Berger R. Acute lymphoblastic leukemia and chromosome 21. Cancer Genetics and Cytogenetics. 1997;94(1):8-12.

67. Cortes JE, Kantarjian H, O'Brien S, Keating M, Pierce S, Freireich EJ, et al. Clinical and prognostic significance of trisomy 21 in adult patients with acute myelogenous leukemia and myelodysplastic syndromes. Leukemia. 1995;9(1):115-7.

68. Segel GB, Lichtman MA. Familial (inherited) leukemia, lymphoma, and myeloma: an overview. Blood Cells, Molecules, and Diseases. 2004;32(1):246-61.

69. Ganmore I, Smooha G, Izraeli S. Constitutional aneuploidy and cancer predisposition. Human Molecular Genetics. 2009;18(R1):R84-93.

70. Martin Belson BK, & Adrianne Holmes. Risk Factors for Acute Leukemia in Children: A Review. Environmental Health Perspectives. 2007;115(1):138-45.

71. Santiago M, Cristina M, Muhterem B, Jingly F, Mona L, Larry M, et al. Preimplantation diagnosis of the aneuploidies most commonly found in spontaneous abortions and live births: XY, 13, 14, 15, 16, 18, 21, 22. Prenatal Diagnosis. 1998;18(13):1459-66.

72. Tallman MS, Neuberg D, Bennett JM, Francois CJ, Paietta E, Wiernik PH, et al. Acute megakaryocytic leukemia: the Eastern Cooperative Oncology Group experience. Blood. 2000;96(7):2405-11.

73. Massey GV, Zipursky A, Chang MN, Doyle JJ, Nasim S, Taub JW, et al. A prospective study of the natural history of transient leukemia (TL) in neonates with

Down syndrome (DS): Children's Oncology Group (COG) study POG-9481. Blood. 2006;107(12):4606-13.

74. Maserati E, Pressato B, Valli R, Patitucci F, Lo Curto F, Pasquali F, et al. Constitutional trisomy 8 mosaicism in primary myelofibrosis: relevance to clinical practice and warning for trisomy 8 studies. Cancer Genetics and Cytogenetics. 2007;179(1):79-81.

75. Macedo Silva, Luiza M, Raimondi SC, Abdelhay E, Gross M, Mkrtchyan H, et al. Banding and molecular cytogenetic studies detected a CBFB-MYH11 fusion gene that appeared as abnormal chromosomes 1 and 16 in a baby with acute myeloid leukemia FAB M4-Eo. Cancer Genetics and Cytogenetics. 2008;182(1):56-60.

76. Stochholm K, Juul S, Juel K, Naeraa RW, Hojbjerg Gravholt C. Prevalence, Incidence, Diagnostic Delay, and Mortality in Turner Syndrome. Journal of Clinical Endocrinology and Metabolism. 2006;91(10):3897-902.

77. Baumgartner BJ, Shurafa M, Terebelo H, Tapazoglou E, Van Dyke DL. Trisomy 15, Sex Chromosome Loss, and Hematological Malignancy. Cancer Genetics and Cytogenetics. 2000;117(2):132-5.

78. Herens C, Brasseur E, Jamar M, Vierset L, Schoenen I, Koulischer L. Loss of the Y chromosome in bone marrow cells: results on 1907 consecutive cases of leukaemia and preleukaemia. Clincal and Laboratory Haematology. 1999;21(1):17-20.

79. Wiktor A, Rybicki BA, Piao ZS, Shurafa M, Barthel B, Maeda K, et al. Clinical significance of Y chromosome loss in hematologic disease. Genes, Chromosomes and Cancer. 2000;27(1):11-6.

80. Lalande M. Parental imprinting and human disease. Annual Review of Genetics. 1996;30(1):173-95.

81. Swerdlow A, Schoemaker M, Higgins C, Wright A, Jacobs PA. Mortality risks in patients with constitutional autosomal chromosome deletions in Britain: a cohort study Human Genetics. 2008;123(2):215-24.

82. Raghavan M, Lillington DM, Skoulakis S, Debernardi S, Chaplin T, Foot NJ, et al. Genome-Wide Single Nucleotide Polymorphism Analysis Reveals Frequent Partial Uniparental Disomy Due to Somatic Recombination in Acute Myeloid Leukemias. Cancer Research. 2005;65(2):375-8.

83. Hall J. Genomic Imprinting: Nature and Clinical Relevance. Annual Review of Medicine. 1997;48(1):35-44.

84. Tharapel AT, Kadandale JS, Martens PR, Wachtel SS, Wilroy RS. Prader Willi/Angelman and DiGeorge/velocardiofacial syndrome deletions: Diagnosis by primed in situ labeling (PRINS). American Journal of Medical Genetics. 2002;107(2):119-22.

85. Clayton-Smith J, Laan L. Angelman syndrome: a review of the clinical and genetic aspects Journal of Medical Genetics. 2003;40:87-95.

86. Gilbert J, Gore SD, Herman JG, Carducci MA. The Clinical Application of Targeting Cancer through Histone Acetylation and Hypomethylation Clinical Cancer Research. 2004;10:4589-96.

87. Baylin SB, Esteller M, Rountree MR, Bachman KE, Schuebel K, Herman JG. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. Human Molecular Genetics. 2001;10(7):687-92.

88. Chong SS, Pack SD, Roschke AV, Tanigami A, Carrozzo R, Smith AC, et al. A revision of the lissencephaly and Miller-Dieker syndrome critical regions in chromosome 17p13.3. Human Molecular Genetics. 1997;6(2):147-55.

89. Claude Caron De F, Thierry S. TP53 tumor suppressor gene: A model for investigating human mutagenesis. Genes, Chromosomes and Cancer. 1992;4(1):1-15.

90. Stocklein H, Smardova J, Macak J, Katzenberger T, Holler S, Wessendorf S, et al. Detailed mapping of chromosome 17p deletions reveals HIC1 as a novel tumor suppressor gene candidate telomeric to TP53 in diffuse large B-cell lymphoma. Oncogene. 2007;27(18):2613-25.

91. Shizuyo U, Masaru K, Shigekazu K, Michihiko W. Gallbladder cancer in a patient with Miller-Dieker syndrome. Acta Paediatrica. 2006;95(1):113-4.

92. Czuchlewski DR, Andrews J, Madden R, Clericuzio CL, Zhang Q-Y. Acute Lymphoblastic Leukemia in a Patient With Miller-Dieker Syndrome. Journal of Pediatric Hematology/Oncology. 2008;30(11):865-8

93. Li FP, Fraumeni JF. Soft tissue sarcomas, breast cancer and other neoplasms: a familial syndrome. Annals of Internal Medicine. 1969;71:747-52.

94. Varley JM. Germline TP53 Mutatiosn and Li-Fraumeni Syndrome. Human Mutation. 2003;21:313-20.

95. Cheung KJJ, Douglas EH, Randy DG. The significance of TP53 in lymphoid malignancies: mutation prevalence, regulation, prognostic impact and potential as a therapeutic target. British Journal of Haematology. 2009;146(3):257-69.

96. Marronet A, Dokal I. Dyskeratosis congenita: a disorder of telomerase deficiency and its relationship to other diseases. Expert review of Dermatology. 2006;1(3):463-79.

97. Fanconi G. Familiaere infantile perniziosaartige Anaemie (pernizioeses Blutbild and konstitution). Jahrbuch Kinderheild. 1927;117:257-80.

98. Wang W. Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. Nature Reviews Genetics. 2007;8(10):735-48.

99. D'Andrea AD. The Fanconi road to cancer. Genes and Development. 2003;17:1933-6.

100. Nishimoto N, Imai Y, Ueda K, Nakagawa M, Shinohara A, Ichikawa M, et al. T cell acute lymphoblastic leukemia arising from familial platelet disorder. International Journal of Hematology. 2010;92(1):194-7.

101. Welte K, Zeidler C. Severe Congenital Neutropenia. Hematology/Oncology Clinics of North America. 2009;23(2):307-20.

102. Carlsson Gr, Andersson M, P√ºtsep K, Garwicz D, Nordenskj√∂ld M, Henter J-I, et al. Kostmann syndrome or infantile genetic agranulocytosis, part one: Celebrating 50 years of clinical and basic research on severe congenital neutropenia. Acta Paediatrica. 2006;95(12):1526 - 32.

103. Kastan MB, Bartek J. Cell-cycle checkpoints and cancer. Nature. 2004;432(7015):316-23.

104. Hartwell LH, Kastan MB. Cell cycle control and cancer. Science. 1994;266(5192):1821-8.

105. Maddika S, Ande SR, Panigrahi S, Paranjothy T, Weglarczyk K, Zuse A, et al. Cell survival, cell death and cell cycle pathways are interconnected: Implications for cancer therapy. Drug Resistance Updates. 2007;10(1-2):13-29.

106. Wotton S, Terry A, Kilbey A, Jenkins A, Herzyk P, Cameron E, et al. Gene array analysis reveals a common Runx transcriptional programme controlling cell adhesion and survival. Oncogene. 2008;27(44):5856-66.

107. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, et al. Recurring Mutations Found by Sequencing an Acute Myeloid Leukemia Genome. New England Journal of Medicine. 2009;361(11):1058-66.

108. Fearon ER. Human Cancer Syndromes: Clues to the Origin and Nature of Cancer. Science. 1997;278(5340):1043-50.

109. Ponder BAJ. Cancer genetics. Nature. 2001;411(6835):336-41.

110. Charlesworth JC, Dyer TD, Stankovich JM, Blangero J, Mackey DA, Craig JE, et al. Linkage to 10q22 for Maximum Intraocular Pressure and 1p32 for Maximum Cup-to-Disc Ratio in an Extended Primary Open-Angle Glaucoma Pedigree. Investigative Ophthalmology and Visual Science. 2005;46(10):3723-9.

111. FitzGerald LM, Patterson B, Thomson R, Polanowski A, Quinn S, Brohede J, et al. Identification of a prostate cancer susceptibility gene on chromosome 5p13q12 associated with risk of both familial and sporadic disease. European Journal of Human Genetics. 2009;17(3):368-77.

112. Lickiss JN, Baikie AG, Panton J. Lymphoproliferative and myeloproliferative disease in Tasmania. The National Cancer Institute Monograph. 1977;47:37-9.

113. Lickiss JN, Giles GG, Baikie MJ, Lowenthal RM, Challis D, Panton J. Myeloproliferative and Lymphoproliferative Disorders in Tasmania, 1972-80: Patterns in Space and Time. The Journal of the National Cancer Institute 1984;72(6):1223-28.

114. McKay JD, Lesueur F, Jonard L, Pastore A, Williamson J, Hoffman L, et al. Localization of a Susceptibility Gene for Familial Nonmedullary Thyroid Carcinoma to Chromosome 2q21. The American Journal of Human Genetics. 2001;69(2):440-6.

115. Rubio JP, Bahlo M, Butzkueven H, van der Mei IAF, Sale MM, Dickinson JL, et al. Genetic Dissection of the Human Leukocyte Antigen Region by Use of Haplotypes of Tasmanians with Multiple Sclerosis. The American Journal of Human Genetics. 2002;70(5):1125-37.

116. Agarwal SK, Kester MB, Debelenko LV, Heppner C, Emmert-Buck MR, Skarulis MC, et al. Germline mutations of the MEN1 gene in familial multiple endocrine neoplasia type 1 and related states. Human Molecular Genetics. 1997;6(7):1169-75.

117. Larsson C, Skogseid B, Oberg K, Nakamura Y, Nordenskjold M. Multiple endocrine neoplasia type 1 gene maps to chromosome 11 and is lost in insulinoma. Nature. 1988;332(6159):85-7.

118. Cazier J-B, Tomlinson I. General lessons from large-scale studies to identify human cancer predisposition genes. The Journal of Pathology. 2010;220(2):255-62.

119. Frank TS. Hereditary Cancer Syndromes. Archives of Pathology & Laboratory Medicine. 2001;125(1):85-90.

120. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science. 1990;250(4988):1684-9.

121. Wooster R, Neuhausen S, Mangion J, Quirk Y, Ford D, Collins N, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science. 1994;265(5181):2088-90.

122. Turnbull C, Rahman N. Genetic Predisposition to Breast Cancer: Past, Present, and Future. Annual Review of Genomics and Human Genetics. 2008;9(1):321-45.

123. Galvan A, Ioannidis JPA, Dragani TA. Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer. Trends in Genetics. 2010;26(3):132-41.

124. Aaltonen L, Peltomaki P, Leach F, Sistonen P, Pylkkanen L, Mecklin J, et al. Clues to the pathogenesis of familial colorectal cancer. Science. 1993;260(5109):812-6.

125. Kastrinos F, Syngal S. Recently Identified Colon Cancer Predispositions: MYH and MSH6 Mutations. Seminars in Oncology. 2007;34(5):418-24.

126. Lynch HT, Shaw MW, Magnuson CW, Larsen AL, Krush AJ. Hereditary Factors in Cancer: Study of Two Large Midwestern Kindreds. Arch Intern Med. 1966;117(2):206-12.

127. Kaelin Jr WG. The von Hippel-Lindau tumour suppressor protein: O2 sensing and cancer. Nat Rev Cancer. 2008;8(11):865-73.

128. Ong KR, Woodward ER, Killick P, Lim C, Macdonald F, Maher ER. Genotype–phenotype correlations in von Hippel-Lindau disease. Human mutation. 2007;28(2):143-9.

129. Nickerson ML, Jaeger E, Shi Y, Durocher JA, Mahurkar S, Zaridze D, et al. Improved Identification of von Hippel-Lindau Gene Alterations in Clear Cell Renal Tumors. Clinical Cancer Research. 2008;14(15):4726-34.

130. Crowther-Swanepoel D, Houlston RS. Genetic variation and risk of chronic lymphocytic leukaemia. Seminars in Cancer Biology. 2010;20(6):363-369.

131. Cartwright RA, Bernard SM, Bird CC, Darwin CM, O'Brien C, Richards ID, et al. Chronic lymphocytic leukaemia: case control epidemiological study in Yorkshire. British Journal of Cancer. 1987;56(1):79-82.

132. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. Systematic Population-Based Assessment of Cancer Risk in First-Degree Relatives of Cancer

Probands. The Journal of the National Cancer Institute 1994 November 2, 1994;86(21):1600-8.

133. Gunz FW, Gunz JP, Chapman CJ, Houston IB. Familial Leukaemia: A Study of 909 Families. Scandinavian Journal of Haematology. 1975;15(2):117-31.

134. Linet M, Van Natta M, Brookmeyer R, Muin J, Mccaffrey L, Humphrey R, et al. Familial Cancer history and chronic lymphocytic leukemia. American Journal of Epidemiology. 1989 November 1, 1989;130(4):655-64.

135. Pottern LM, Linet M, Blair A, Dick F, Burmeister LF, Gibson R, et al. Familial cancers associated with subtypes of leukemia and non-hodgkin's lymphoma. Leukemia Research. 1991;15(5):305-14.

136. Fuller SJ, Papaemmanuil E, McKinnon L, Webb E, Sellick GS, Dao-Ung L, et al. Analysis of a large multi-generational family provides insight into the genetics of chronic lymphocytic leukemia. British Journal of Haematology. 2008 142: 238-45.

137. Ng D, Toure O, Wei M-H, Arthur DC, Abbasi F, Fontaine L, et al. Identification of a novel chromosome region, 13q21.33-q22.2, for susceptibility genes in familial chronic lymphocytic leukemia. Blood. 2007;109(3):916-25.

138. Sellick GS, Webb EL, Allinson R, Matutes E, Dyer MJS, Jansson V, et al. A High-Density SNP Genomewide Linkage Scan for Chronic Lymphocytic Leukemia Susceptibility Loci. American journal of human genetics. 2005;77(3):420-9.

139. McInnis MG. Anticipation: An old idea in new genes. American Journal of Human Genetics. 1996;59:973-9.

140. Daugherty SE, Pfeiffe RM, Mellemkjaer L, Hemminki K, Goldin LR. No Evidence for Anticipation in Lymphoproliferative Tumors in Population-Based Samples Cancer Epidemiology, Biomarkers and Prevention 2005;14:1245-50.

141. Pearson CE, Edamura KN, Cleary JD. Repeat instability: mechanisms of dynamic mutations. Nature Reviews Genetics. 2005;6(10):729-42.

142. Benzow KA, Koob MD, Condie A, Catovsky D, Matutes E, Yuille MR, et al. Instability of CAG-trinucleotide Repeats in Chronic Lymphocytic Leukemia. Leukemia and Lymphoma. 2002;43(10):1987 - 90.

143. Klein RJ, Xu X, Mukherjee S, Willis J, Hayes J. Successes of Genome-wide Association Studies. Cell. 2010;142(3):350-1.

144. Crowther-Swanepoel D, Broderick P, Di Bernardo MC, Dobbins SE, Torres M, Mansouri M, et al. Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. Nat Genet. 2010;42(2):132-6.

145. Papaemmanuil E, Hosking FJ, Vijayakrishnan J, Price A, Olver B, Sheridan E, et al. Loci on 7p12.2, 10q21.2 and 14q11.2 are associated with risk of childhood acute lymphoblastic leukemia. Nature Genetics. 2009;41(9):1006-10.

146. Skibola CF, Bracci PM, Halperin E, Conde L, Craig DW, Agana L, et al. Genetic variants at 6p21.33 are associated with susceptibility to follicular lymphoma. Nat Genet. 2009;41(8):873-5.

147. Slager SL, Goldin LR, Strom SS, Lanasa MC, Spector LG, Rassenti L, et al. Genetic Susceptibility Variants for Chronic Lymphocytic Leukemia. Cancer Epidemiology Biomarkers & Prevention. 2010;19(4):1098-102.

148. Trevino LR, Yang W, French D, Hunger SP, Carroll WL, Devidas M, et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. Nature Genetics. 2009;41(9):1001-7.

149. Di Bernardo MC, Crowther-Swanepoel D, Broderick P, Webb E, Sellick G, Wild R, et al. A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. Nature Genetics. 2008;40(10):1204-10.

150. Rudd MF, Sellick GS, Webb EL, Catovsky D, Houlston RS. Variants in the ATM-BRCA2-CHEK2 axis predispose to chronic lymphocytic leukemia. Blood. 2006;108(2):638-44.

151. Trevino LR, Yang W, French D, Hunger SP, Carroll WL, Devidas M, et al. Germline genomic variants associated with childhood acute lymphoblastic leukemia. Nat Genet. 2009;41(9):1001-5.

152. Yang W, Trevino LR, Yang JJ, Scheet P, Pui CH, Evans WE, et al. ARID5B SNP rs10821936 is associated with risk of childhood acute lymphoblastic leukemia in blacks and contributes to racial differences in leukemia incidence. Leukemia. 2010;24:894-896.

153. Leone G, Mele L, Pulsoni A, Equitani F, Pagano L. The incidence of secondary leukemias. Haematologica. 1999;84(10):937-45.

154. Knight JA, Skol AD, Shinde A, Hastings D, Walgren RA, Shao J, et al. Genome-wide association study to identify novel loci associated with therapy-related myeloid leukemia susceptibility. Blood. 2009;113(22):5575-82.

155. Mardis ER. Next-Generation DNA Sequencing Methods. Annual Review of Genomics and Human Genetics. 2008;9(1):387-402.

156. Meyerson M, Gabriel S, Getz G. Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet. 2010;11(10):685-96.

157. von Bubnoff A. Next-Generation Sequencing: The Race Is On. Cell. 2008;132(5):721-3.

158. Werner T. Next generation sequencing in functional genomics. Briefings in Bioinformatics. 2010;11(5):499-511.

159. Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, et al. The complete genome of an individual by massively parallel DNA sequencing. Nature. 2008;452(7189):872-6.

160. Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature. 2008;456(7218):66-72.

161. ISCN 2009 an international system for human cytogenetic nomenclature Shaffer L, Slovak M, Campbell L, editors: Springer, Berlin/Heidelberg; 2009.

162. Tegg EM, Thomson RJ, Stankovich JM, Banks A, Marsden KA, Lowenthal RM, et al. Anticipation in familial hematological malignancies. Blood. 2011;117(4):1308-10.

163. Bittles AH. Consanguinity and its relevance to clinical genetics. Clinical Genetics. 2001;60(2):89-98.

164. Modell B, Darr A. Genetic counselling and customary consanguineous marriage. Nature Reviews Genetics. 2002;3(3):225-9.

165. Lister A, Abrey LE, Sandlund JT. Central Nervous System Lymphoma. Hematology. 2002(1):283-96.

166. Akintola O, Akintola-Ogunremi O, Whitney, Whitney C, Mathur, Mathur S, et al. Chronic lymphocytic leukemia presenting with symptomatic central nervous system involvement. Annals of Hematology. 2002;81(7):402-4.

167. Tegg EM, Thomson RJ, Stankovich J, Banks A, Flowers C, McWhirter R, et al. Evidence for a common genetic aetiology in high-risk families with multiple haematological malignancy subtypes. British Journal of Haematology. 2010;150(4):456-62.

168. Abed R, Bourdon V, Huiart L, Eisinger F, Khelif A, Frenay M, et al. Molecular study of CEPBA in familial hematological malignancies. Familial Cancer. 2009;8(4):581-4.

169. Discoveries in neuroscience. Annals of Neurology. 2010;67(1):A16-A.

170. Lynch HT, Weisenburger DD, Quinn-Laquer B, Watson P, Lynch JF, Sanger WG. Hereditary chronic lymphocytic leukemia: An extended family study and literature review. American Journal of Medical Genetics. 2002;115(3):113-7.

171. Rawstron AC, Yuille MR, Fuller J, Cullen M, Kennedy B, Richards SJ, et al. Inherited predisposition to CLL is detectable as subclinical monoclonal B-lymphocyte expansion. Blood. 2002;100(7):2289-90.

172. Gunnarsson R, Staaf J, Jansson M, Ottesen AM, Gransson H, Ulrika. L, et al. Screening for copy-number alterations and loss of heterozygosity in chronic

lymphocytic leukemia-A comparative study of four differently designed, high resolution microarray platforms. Genes, Chromosomes and Cancer. 2008;47(8):697-711.

173. Houlston RS, Sellick G, Yuille M, Matutes E, Catovsky D. Causation of chronic lymphocytic leukemia--insights from familial disease. Leukemia Research. 2003;27(10):871-6.

174. Schmidt HH, Dyomin VG, Palanisamy N, Itoyama T, Nanjangud G, Pirc-Danoewinata H, et al. Deregulation of the carbohydrate (chondroitin 4) sulfotransferase 11 (CHST11) gene in a B-cell chronic lymphocytic leukemia with a t(12;14)(q23;q32). Oncogene. 2004;23(41):6991-6.

175. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proceedings of the National Academy of Sciences of the United States of America. 2002;99(24):15524-9.

176. Gerald M, Fatima A, Elizabeth R, Andy CR, Paolo G, Therese A, et al. Overview of monoclonal B-cell lymphocytosis. British Journal of Haematology. 2007;139(5):701-8.

177. Tute R, Yuille M, Catovsky D, Houlston RS, Hillmen P, Rawstron AC. Monoclonal B-cell lymphocytosis (MBL) in CLL families: substantial increase in relative risk for young adults. Leukemia. 2006;20:728-729.

178. Isidre F, Judith A, Sabina C, Piero P, Thomas A, Jeanne B, et al. Effects of Formalin Fixation, Paraffin Embedding, and Time of Storage on DNA Preservation in Brain Tissue: A BrainNet Europe Study. Brain Pathology. 2007;17(3):297-303.

179. Ananian V, Tozzo P, Ponzano E, Nitti D, Rodriguez D, Caenazzo L. Tumoural specimens for forensic purposes: comparison of genetic alterations in frozen and formalin-fixed paraffin-embedded tissues. International Journal of Legal Medicine.

180. Weise A, Timmermann B, Grabherr M, Werber M, Heyn P, Kosyakova N, et al. High-throughput sequencing of microdissected chromosomal regions. European Journal of Human Genetics. 2009;18(4):457-62.

181. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, et al. Direct selection of human genomic loci by microarray hybridization. Nature Methods. 2007;4(11):903-5.

182. Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, et al. Genome-wide in situ exon capture for selective resequencing. Nature Genetics. 2007;39(12):1522-7.

183. Pinard R, de Winter A, Sarkis G, Gerstein M, Tartaro K, Plant R, et al. Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. BMC Genomics. 2006;7(1):216.

184. Jonsson V, Houlston RS, Catovsky D, Yuille MR, Hilden J, Olsen JH, et al. CLL family Pedigree 14 revisited: 1947-2004. Leukemia. 2005;19(6):1025-8.

185. Han S, Kyoung-Mu M, Lee A, Park S, Lee J, Ahn H, et al. Genome-wide association study of childhood acute lymphoblastic leukemia in Korea. Leukemia Research. 2010;34(10):1271-4.

186. Skibola CF, Curry JD, Nieters A. Genetic susceptibility to lymphoma. Haematologica. 2007;92(7):960-9.

187. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nature Genetics. 2004;36(9):949-51.

188. Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, et al. Mapping and sequencing of structural variation from eight human genomes. Nature. 2008;453(7191):56-64.

189. Lee C, Iafrate AJ, Brothman AR. Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. Nature Genetics. 2007.

190. Perry GH, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, Lee AS, et al. Copy number variation and evolution in humans and chimpanzees. Genome Research. 2008;18(11):1698-710.

191. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006;444:444-54.

192. Pfeifer D, Pantic M, Skatulla I, Rawluk J, Kreutz C, Martens UM, et al. Genome-wide analysis of DNA copy number changes and LOH in CLL using high-density SNP arrays. Blood. 2007;109(3):1202-10.

193. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. American journal of human genetics. 2007;81(3):559-75.

194. Kruglyak L, Daly MJ, Reeve-Daly MP, ES L. Parametric and nonparametric linkage analysis: a unified multipoint approach. American Journal of Human Genetics. 1996;58(6):1347-63.

195. Dyke B. PEDSYS, a Pedigree Data Management System User's Manual: Population Genetics Laboratory Technical Report No. 2. . 2nd ed. San Antonio, TX 78227:: Southwest Foundation for Biomedical Research; 1993.

196. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, et al. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Research. 2007;17(11):1665-74.

197. Rabbee N, Speed TP. A genotype calling algorithm for affymetrix SNP arrays. Bioinformatics.2005;22(1):7-12.

198. Cuneo A, Roberti MG, Bigoni R, Minotto C, Bardi A, Milani R, et al. Four novel non-random chromosome rearrangements in B-cell chronic lymphocytic leukaemia: 6p24-25 and 12p12-13 translocations, 4q21 anomalies and monosomy 21. British Journal of Haematology. 2000;108(3):559-64.

199. Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(39):13944-9.

200. Vincent M, Olivier E, Mario H, Chantal G, …ric L. Copy-number variations (CNVs) of the human sex steroid metabolizing genes UGT2B17 and UGT2B28 and their associations with a UGT2B15 functional polymorphism. Human mutation. 2009;30(9):1310-9.

201. Xue Y, Sun D, Daly A, Yang F, Zhou X, Zhao M, et al. Adaptive Evolution of UGT2B17 Copy-Number Variation. The American Journal of Human Genetics. 2008;83(3):337-46.

202. Yang T-L, Chen X-D, Guo Y, Lei S-F, Wang J-T, Zhou Q, et al. Genome-wide Copy-Number-Variation Study Identified a Susceptibility Gene, UGT2B17, for Osteoporosis. The American Journal of Human Genetics. 2008;83(6):663-74.

203. McCarroll SA, Bradner JE, Turpeinen H, Volin L, Martin PJ, Chilewski SD, et al. Donor-recipient mismatch for common gene deletion polymorphisms in graft-versus-host disease. Nature Genetics. 2009;41(12):1341-4.

204. Houlston RS, Catovsky D, Yuille MR. Genetic susceptibility to chronic lymphocytic leukemia. Leukemia. 2002;16(6):1008-14.

205. Vierimaa O, Georgitsi M, Lehtonen R, Vahteristo P, Kokko A, Raitila A, et al. Pituitary Adenoma Predisposition Caused by Germline Mutations in the AIP Gene. Science. 2006;312(5777):1228-30.

206. de la Chapelle A. Genetic predisposition to human disease: allele-specific expression and low-penetrance regulatory loci. Oncogene. 2009;28(38):3345-8.

207. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nature Biotechnology. 1996;14(13):1675-80.

208. Hardiman G. Microarray platforms - comparisons and contrasts. Pharmacogenomics. 2004;5(5):487-502.

209. Xing Y, Kapur K, Wong WH. Probe Selection and Expression Index Computation of Affymetrix Exon Arrays. PLoS ONE. 2006;1(1):e88.

210. Affymetric I. Exon_array_design_technote. Affymetric; 2005.

211. Neitzel H. A routine method for the establishment of permanent growing lymphoblastoid cell lines. Human Genetics. 1986;73(4):320-6.

212. Carter KL, Cahir-McFarland E, Kieff E. Epstein-Barr Virus-Induced Changes in B-Lymphocyte Gene Expression. The Journal of Virology. 2002 October 15, 2002;76(20):10427-36.

213. Wren C, Moriarty H, Marsden K, Tegg E. Cytogenetic investigations of chronic lymphocytic leukemia. Cancer Genetics and Cytogenetics. 2010;198(2):155-61.

214. Yeoh E-J, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. 2002;1(2):133-43.

215. Armstrong SA, Staunton JE, Silverman LB, Pieters R, den Boer ML, Minden MD, et al. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nature Genetics. 2002;30(1):41-7.

216. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000;403(6769):503-11.

217. Chim C, Fung T, Wong K, Lau J, Liang R. Frequent DAP kinase but not p14 or Apaf-1 hypermethylation in B-cell chronic lymphocytic leukemia. Journal of Human Genetics. 2006;51(9):832-8.

218. Friedman LM, Dror AA, Mor E, Tenne T, Toren G, Satoh T, et al. MicroRNAs are essential for development and function of inner ear hair cells in vertebrates. Proceedings of the National Academy of Sciences. 2009 May 12, 2009;106(19):7915-20.

219. Sarkar A, Sim C, Hong Y, Hogan J, Fraser M, Robertson H, et al. Molecular evolutionary analysis of the widespread piggyBac transposon family and related "domesticated" sequences. Molecular Genetics and Genomics. 2003;270(2):173-80.

220. Sinzelle L, Izsvák Z, Ivics Z. Molecular domestication of transposable elements: From detrimental parasites to useful host genes. Cellular and Molecular Life Sciences. 2009;66(6):1073-93.

221. Enjuanes A, Benavente Y, Bosch F, Martin-Guerrero I, Colomer D, Perez-Alvarez S, et al. Genetic Variants in Apoptosis and Immunoregulation-Related Genes Are Associated with Risk of Chronic Lymphocytic Leukemia. Cancer Research. 2008;68(24):10178-86.

222. Levy L, Howell M, Das D, Harkin S, Episkopou V, Hill CS. Arkadia Activates Smad3/Smad4-Dependent Transcription by Triggering Signal-Induced SnoN Degradation. Molecular and Cellular Biology. 2007;27(17):6068-83.

223. Jang C-W, Chen C-H, Chen C-C, Chen J-y, Su Y-H, Chen R-H. TGF-beta induces apoptosis through Smad-mediated expression of DAP-kinase. Nature Cell Biology. 2002;4(1):51-8.

224. Lagneaux L, Delforge A, Bron D, Massy M, Bernier M, Stryckmans P. Heterogenous response of B lymphocytes to transforming growth factor-beta in B-cell chronic lymphocytic leukaemia: correlation with the expression of TGF-B; receptors. British Journal of Haematology. 1997;97(3):612-20.

225. DeCoteau JF, Knaus PI, Yankelev H, Reis MD, Lowsky R, Lodish HF, et al. Loss of functional cell surface transforming growth factor B (TGF-B) type 1 receptor correlates with insensitivity to TGF-B in chronic lymphocytic leukemia. Proceedings of the National Academy of Sciences of the United States of America. 1997;94(11):5877-81.

226. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. The Diploid Genome Sequence of an Individual Human. PLoS Biol. 2007;5(10):e254.

227. von Winiwarter. H. Études sur la spermatogenese humaine. . Arch biologie. 1912; 27(93):147-9.

228. Tjio J.H., Levan A. The chromsome number of man. Hereditas. 1956;42(1-2):1-6.

229. Watson JD, Crick FHC. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature. 1953;171(4356):737-8.

230. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, et al. Nucleotide sequence of bacteriophage phi X174 DNA. Nature. 1977;265(Feb):687-95.

231. Seeburg PH SJ, Martial JA, Ullrich A, Goodman HM, Baxter JD. Nucleotide sequence of a human gene coding for a polypeptide hormone. Transactions of the Association of American Physicians. 1977;90:109-16.

232. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The Sequence of the Human Genome. Science. 2001;291(5507):1304-51.

233. Schlenk RF, Dohner K, Krauter Jr, Frohling S, Corbacioglu A, Bullinger L, et al. Mutations and Treatment Outcome in Cytogenetically Normal Acute Myeloid Leukemia. New England Journal of Medicine. 2008;358(18):1909-18.

234. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-73.

235. Biesecker LG. Exome sequencing makes medical genomics a reality. Nat Genet. 2010;42(1):13-4.

236. Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet. 2010;42(9):790-3.

237. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, et al. DNMT3A Mutations in Acute Myeloid Leukemia. New England Journal of Medicine. 2010; 363;2424-2433.

238. Zhou J, Wang H, Lu A, Hu G, Luo A, Ding F, et al. A novel gene, NMES1, downregulated in human esophageal squamous cell carcinoma. International Journal of Cancer. 2002;101(4):311-6.